

**A small RNA species from mouse erythroleukaemic
cells is partially homologous to an interspersed middle
repetitive sequence in the mouse genome**

INAUGURAL - DISSERTATION
zur Erlangung
der Philosophischen Doktorwürde

vorgelegt der
Philosophischen Fakultät II
der Universität Zürich

von
CHRISTINE GRÜBER - STEINITZ
von Vugelles - La - Mothe (VD)

1983

begutachtet von Herrn Prof. Dr. C. Weissmann

aku-Fotodruck
Zürich 1982

**A small RNA species from mouse erythroleukaemic
cells is partially homologous to an interspersed middle
repetitive sequence in the mouse genome**

INAUGURAL - DISSERTATION

zur Erlangung

der Philosophischen Doktorwürde

vorgelegt der

Philosophischen Fakultät II

der Universität Zürich

von

CHRISTINE GRÜBER - STEINITZ

von Vugelles - La - Mothe (VD)

begutachtet von Herrn Prof. Dr. C. Weissmann

aku-Fotodruck

Zürich 1982



à Philippe

à mes parents

Acknowledgements

I would like to express my gratitude to Prof. Dr. C. Weissmann, who has introduced me to this field and given me continuous advice and support.

I also wish to thank Dr. P. Curtis who helped me a lot during my first steps in laboratory work and Dr. P.D. H. Weber for many valuable discussions and critical reading of this manuscript.

I am also grateful to Prof. Dr. M. Billeter for helpful discussions on RNA sequencing.

I wish to thank all my previous and actual collaborators in the institute for scientific advice, technical help and moral support.

Without the help of Connie Dierks, who did the typing, this thesis would not be here.



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41507210_0006

Contents

Summary

1. Introduction

1.1. mRNA precursors in eucaryotic cells

1.2. Split genes and RNA splicing

1.3. The origin of this work

2. Materials and Methods

2.1. Materials

2.2. Methods

2.2.1. Tissue culture

2.2.2. Total cell RNA isolation

2.2.3. RNA sequence determination

2.2.3.1. RNase T_1 and RNase A fingerprints

2.2.3.2. Sequence determination of the RNase T_1
and RNase A oligonucleotides

2.2.4. Restriction analysis, Southern hybridization
and DNA sequence determination

3. Results

3.1. Discovery of RNA 16 in Friend cells

3.2. Yields of (^{32}P) RNA 16 from Friend cells under various
labelling conditions

3.3. Comparison of the various small RNA species by finger-
print analysis

3.4. Partial sequence of RNA 16

3.4.1. RNase T_1 and RNase A fingerprints

3.4.2. Catalogues of the RNase T_1 and RNase A oligo-
nucleotides

- 3.4.3. Comparison of the data with the complete sequence of RNA 16 published by Harada and Kato (1980)
- 3.5. Occurrence and intracellular distribution of RNA 16
 - 3.5.1. Correlation of the appearance of RNA 16 with cell induction
 - 3.5.2. Intracellular location of RNA 16
 - 3.5.3. Presence of RNA 16 in mouse L-cells
 - 3.5.4. Association of RNA 16 with high molecular weight poly(A) RNA
 - 3.5.5. Occurrence of RNA 16 in monomeric RNP particles
- 3.6. Relationship of RNA 16 to Z-pBR322/Mchr β G-36 DNA
 - 3.6.1. Restriction analysis of the DNA beyond the 3' end of the β -globin gene
 - 3.6.2. Localization on Z-pBR322/Mchr β G-36 DNA of the sequence related to RNA 16 by Southern hybridization experiments
 - 3.6.3. Frequency of sequences related to the AluI/BspI 262 bp fragment in the total mouse genome
 - 3.6.4. Identification of the strand hybridizing to RNA 16
 - 3.6.5. Determination of the nucleotide sequence of the region of plasmid Z-pBR322/Mchr β G-36 involved in binding RNA 16
- 3.7. Structure of the DNA - RNA hybrid
 - 3.7.1. RNase T₁ treatment of the DNA - RNA hybrid
 - 3.7.2. Partial degradation by nuclease S1 of the hybrid using terminally (³²P)-labeled DNA

3.7.3. Footprinting experiment using sodium bisulfite
as a single-strand specific reagent

4.1 Discussion

- 4.1. The nucleotide sequence of RNA 16 and its comparison to other known small RNA molecules
- 4.2. Intracellular distribution of small RNA molecules and speculations on their possible role
- 4.3. Association of RNA 16 with high molecular-weight poly(A) RNA
- 4.4. Association of low molecular weight RNA with RNP particles
- 4.5. Relationship between RNA 16 and the other RNA species hybridizing to Z-pBR322/Mchr8G-36
- 4.6. What is the role of Alu sequences in the eucaryotic genome?

References

Appendix

- Details of the sequence determination of the RNase T₁ and RNase A oligonucleotides of RNA 16 by further enzyme digestion
- Possible secondary structures of RNA 16 and of the single-stranded DNA binding RNA 16
- Hypothetical models of the DNA-RNA hybrid

Summary

A small RNA species was isolated from induced and non-induced Friend erythroleukaemic cells as well as mouse L-cells. It was found both in nuclei and cytoplasm and shown to be associated with high molecular weight poly(A) RNA. It is also suspected to be present in RNP particles. Upon electrophoresis on PAA gels the small RNA species was separated into 4 distinct spots, each spot corresponding to an RNA molecule with a length of 91 - 94 bases respectively. As shown by partial sequence analysis, this heterogeneity is due to a variable number of U residues at the 3' end. The 5' terminus is pppG.. and the content of G is almost 40%. The small RNA hybridizes to the mouse chromosomal β -globin DNA segment of Tilghman et al. (1977), to a region 2300 - 2570 bases downstream from the 3' end of the globin gene. It exhibits an identical fingerprint with RNA 16 isolated by Peters et al. (1977) from Moloney Murine Leukaemia virus infected 3T3 cells. An RNA species with similar properties and the same fingerprint was found in murine cells by Harada and Kato (1980), who determined its complete nucleotide sequence. After comparison of this sequence with the globin DNA, we were able to show that the hybridizing DNA region is only partially homologous to the RNA. Characterization of the base-paired regions of the RNA - DNA hybrid by several types of footprinting experiments yielded results which are in partial agreement with the interaction predicted from the nucleotide sequences. A Southern experiment with a nick translated DNA fragment carrying the partial homology to RNA 16 as a probe indicated a high frequency of related sequences in the total mouse genome. Both the small RNA and the DNA segment on

the globin plasmid contain a sequence characteristic for the "Alu family" of interspersed repeated sequences found in mammalian genomes (Houck et al., 1979). Possible functions for RNA 16 and its potential relationship to Alu sequences in mammalian cells are discussed.

1. INTRODUCTION

1.1. mRNA precursors in eukaryotic cells

In bacteria the messenger RNA is a primary transcript. Even as the mRNA molecule is being transcribed, it is bound to ribosomes and protein synthesis is initiated. In eukaryotic cells, however, the primary transcript of the gene is not used directly as a messenger RNA. Rather, it undergoes several post-transcriptional modifications, such as methylation, 5' terminal capping, 3' terminal shortening, polyadenylation and endonucleolytic cleavages followed by religation of specific fragments (splicing). A major aim in the past few years has been to determine the possible mechanisms responsible for mature mRNA production as well as the relationship between mRNA and a class of molecules called hnRNA, thought to contain the precursors for eucaryotic messengers. The pioneer work on large RNA precursors goes back to the early sixties, with the discovery of 45S pre-ribosomal RNA molecules. Realization that RNA synthesized in the nucleolus might be a precursor to ribosomal RNA came initially as a result of studies designed to clarify the relationship between nuclear and cytoplasmic RNA (Perry, 1962; Perry, 1964). After a 30' pulse of tritiated nucleoside which results in almost exclusive labelling of nuclear RNA, sedimentation analysis showed label in a polydisperse fraction in the range of 32 S to 45 S, a peak in the 4 S component and lesser amounts in RNA of intermediate sedimentation rates. Shortly after, Scherrer and co-workers (1963) were able to show by "actinomycin-chase" kinetics the precursor-product relationship between 45 S, 32 S, 28 S and 18 S ribosomal RNA species. In these experiments cells

were pulse-labeled with radioactive ribonucleosides and then chased with unlabeled precursors in the presence of actinomycin D, at a level sufficient to inhibit subsequent ribosomal RNA synthesis completely. Sequential labelling of 45 S, followed by 32 - 35 S, 18 S and finally 28 and 18 S components was observed. Base composition analysis, showing that 45 S RNA resembles a composite of the 28 S and 18 S components, and competition hybridisation experiments between 45 S and mature rRNA components indicating a high degree of sequence homologies (Perry et al., 1964) supported the proposed ribosomal RNA maturation events. However, the possibility that the apparent molecular weight of the "ribosomal precursors" was due to artificial aggregation of smaller RNA species, resulting from purification procedures based on hot phenol extraction, was still suspected. The covalent structure of the molecules was subsequently demonstrated (Granboulan and Scherrer, 1969) by exposing the RNA for short times to pH 3 before measuring the length of the molecules in the electron microscope. Recently recombinant DNA techniques have allowed closer investigations of ribosomal RNA genes and their precursors and have supported and completed the data from these early years. The discovery of what we now call heterogeneous nuclear RNA and the establishment of its relationship to messenger RNA was somewhat more tedious. In 1948, Bergstrand et al. had been able to show that cells exposed to radioactive RNA precursors first incorporated the label into the nuclear RNA. Several years later, Goldstein and Plaut (1955) reported that in cells previously labeled for a short time and subsequently transferred to non-radioactive medium, the radioactivity left

the nuclear RNA and appeared in the cytoplasmic RNA fraction. Therefore they assumed that the rapidly labeled RNA in the nucleus was the precursor to the more slowly labeled RNA in the cytoplasm. This idea was first questioned in 1959, when it was observed that there were major discrepancies between the amount of radioactivity lost from the nuclear RNA and the amount that appeared in the cytoplasmic RNA (Watts and Harris, 1959). At that time, however, it was generally assumed, although no specific experiments were done to test the idea, that the information for protein synthesis resided in the RNA of the ribosomes. Since these particles contain more than 85% of the RNA in the cytoplasm it was assumed that the bulk of cytoplasmic RNA, namely ribosomal RNA, would have to fulfill the role of a template for protein synthesis. It was only with the advent of the operator model of Jacob and Monod (1961), that the presence of a new type of RNA, namely that of a short-lived template was postulated. Ribosomal RNA, however, could not fulfill this role since it was known to be very stable (Hershey, 1954). Stimulated by the ideas of Jacob and Monod, Brenner et al. (1961) offered evidence for the existence of messenger RNA in T_4 bacteriophage infected *E.coli*. They found that after infection of *E.coli* with bacteriophage T_4 , no new ribosomes were formed for the synthesis of viral protein, but instead, a new RNA species with a rapid turnover and a base composition complementary to that of phage DNA was produced and attached to preexisting *E.coli* ribosomes. Further support for the mRNA concept was provided by the classic in vitro experiments of Nirenberg and Matthai (1961) who, using

a cell free protein synthesizing system from *E. coli*, showed that addition of synthetic polyuridylic acid resulted in the formation of polyphenylalanine. Progress in the understanding of gene expression in eucaryotes was slower. A defined, eucaryotic (non viral) mRNA was first found in the cytoplasm of reticulocytes by Marbaix et al. (1966) and Lockard and Lingrel (1969) showed that this minor RNA component sedimenting in the 9 S region caused the formation of globin in a cell free protein synthesizing system. The precursor to messenger RNA in eucaryotic cells, later called heterogeneous nuclear RNA because of its size heterogeneity, was found by Scherrer et al. (1963) as a minor component cosedimenting with preribosomal RNA in growing Hela cells. Subsequently, these giant RNA molecules (30 - 100 S) were observed to represent the major component of nuclear RNA in non-dividing cells such as avian erythroblasts (Attardi et al., 1966; Scherrer et al., 1966) or of rapidly labeled RNA in Hela cells (Warner et al., 1966). It appeared likely that at least some of this RNA was a precursor to cytoplasmic messenger RNA, but a direct precursor-product relationship could not be established experimentally as in the case of ribosomal RNA. In addition, the possibility of aggregation resulting from extraction methods or exceptional sedimentation properties due to secondary structure could not be completely ruled out. In 1969, Granboulan and Scherrer undertook to compare published data obtained by more classical methods in order to test the

value of length measurements of RNA molecules in the electron microscope for molecular weight determination. Molecules, whose length had been estimated by a method based on partial hydrolysis (E.coli rRNA: Midgley, 1965; R 17 phage RNA: Sinha et al., 1965) served as a reference for the determination of the average length per nucleotide. In order to disrupt secondary structure, RNA samples were diluted in 8 M urea and spread according to a modification of the protein monolayer technique of Kleinschmidt et al. (1962). Using this method they claimed that the gradient fractions corresponding to sedimentation constants of 80 S contained individual linear molecules up to 30'000 bases in length rather than artificial aggregates or smaller molecules in some particular physical configuration. Once again this view was challenged when several laboratories discovered that strongly denaturing conditions produced considerable shifts in the sedimentation profiles of their RNA preparations which had been previously centrifuged through a non-denaturing sucrose gradient containing high ionic strength buffers (MacNaughton et al., 1974). The explanation for these conflicting findings was given by Federoff et al. (1977) who showed that artefactual aggregation resulting from treatment of the RNA preparation at high ionic strength or by a two phase phenol-aqueous system may be dissolved in presence of strong denaturing agents. This association of RNA sequences was shown upon visualization in the electron microscope to be due to the existence of self-complementary inverted repeats

occurring at a high frequency in hnRNA populations. These sequences would allow the formation of inter- and intramolecular duplexes under certain experimental conditions. Finally, in their communication describing an electrophoretic method for molecular weight determination of RNA molecules, Spohr et al. (1976) conclude that even in presence of a denaturing agent as for example formamide, the RNA preparation has to be electrophoresed at an elevated temperature in order to exclude any possible aggregation effect. To conclude this debate about the actual size distribution of hnRNA molecules, Perry et al. (1975) have been able to extract rapidly labeled RNA molecules in a size range of 0.5 - 30 kb with a mass average of 13 kb from mouse L-cells by using very gentle extraction procedures that minimize degradation and by exposing the RNA to stringent denaturation conditions that avoids artefactual aggregations (heating for 2 min. in Me_2SO /dimethylformamide/1mM EDTA 1:2:1). Several groups have attempted to examine the possibility of a direct precursor-product relationship between the large nuclear RNA molecules and cytoplasmic messenger RNA. In 1970 Scherrer et al. were able to show by means of high Croft competition-hybridization experiments, the presence of common sequences between Hela cell hnRNA and polysomal mRNA. Imaizumi et al. (1973) have reported that globin-specific sequences could be detected in duck erythroblast high molecular weight nuclear RNA. On the other hand, Spohr et al. (1976), using fully denaturing gels, were not able to detect any large molecular weight globin specific RNA. However, these studies were performed with unlabeled

RNA and the cDNA probes available were not pure enough to give a clearcut result. The initial indication for the presence of a precursor to messenger RNA came with the discovery of specific precursor molecules for hemoglobin messengers. In 1974, MacNaughton et al., using cDNA prepared from purified duck hemoglobin messenger RNA as a hybridization probe claimed the presence of hemoglobin mRNA specific sequences sedimenting at 14 S in nuclear RNA of duck erythroblasts. Subsequently, several groups demonstrated the existence of a 15 S globin specific RNA in a variety of erythroid cells such as mouse fetal liver (Ross, 1976), Friend erythroleukaemic (Curtis and Weissmann, 1976; Bastos and Aviv, 1977) and anaemic spleen cells (Kwan et al., 1977). The first clear demonstration of the existence of a globin precursor was given in 1978 by Curtis et al. They purified a 15 S species specific for β -globin messenger RNA either by hybridization of labeled RNA to poly(dC) elongated globin cDNA followed by binding of the hybrid to poly(I) Sephadex or by hybridization to a cloned globin cDNA hybrid plasmid attached to nitrocellulose filters. By T_1 oligonucleotide fingerprint analysis they showed that the 15 S globin RNA contains all characteristic β -globin specific oligonucleotides as well as a defined set of additional nucleotides.

1.2. Split genes and RNA splicing

With the advent of molecular cloning, which made it possible to obtain large quantities of pure DNA fragments the size of a gene, several groups undertook to study the structure and expression of eucaryotic genes. The discovery of discontinuous genes coding for polypeptides however, came from work on the

mRNA of animal viruses, mainly Adenovirus and SV40 (Ribosomal genes of *Drosophila melanogaster* were the first eucaryotic genes described to contain intervening sequences (Glover and Hogness, 1977), however such ribosomal genes are unusual and were found to be rarely transcribed (Long and Dawid, 1979)). In experiments involving hybridization of Adenovirus late mRNAs to DNA fragments derived from restriction enzyme digests, it was found that the 5' end of several viral mRNAs is complementary to regions of DNA some distance away from the area coding for the main body of the RNA (Klessig, 1977; Berget et al., 1977), indicating that segments transcribed from different parts of the DNA were ligated to form the final message. Similar observations were made for SV40 by Celma et al. (1977) and Aloni et al. (1977). At the same time, several independent studies on discontinuous genes of eucaryotic organisms were published. Jeffreys and Flavell (1977) analyzed the structure of the rabbit β -globin gene and surrounding regions by restriction mapping and compared it to the cleavage pattern of β -globin cDNA from Efstratiadis et al. (1977). In the course of this analysis they found that the gene is not a single continuous stretch of DNA but instead is present in at least 2 pieces which are separated by 600 bp non- β -globin sequences. Similar discontinuities were found in the case of the ovalbumin (Breathnach et al., 1977) immunoglobulin (Brack and Tonegawa, 1977) and mouse β -globin genes (Tilghman et al., 1978; Konkel et al., 1978; Tiemeier et al., 1978). Initially it was not clear whether this discontinuity was present in functional genes or whether it was somehow related to inactivation of genes in tissues that

do not express them. Along this line, Jeffreys and Flavell (1977) have shown by restriction analysis of DNA from various rabbit tissues that the β -globin gene contains intervening sequences even in germline cells and in cells actively expressing globin. Similarly, Breathnach et al. (1977) found that the ovalbumin gene is discontinuous in oviduct cells which are actively synthesizing ovalbumin. 1978 Tilghman et al. have shown that the precursor RNA for mouse β -globin forms a continuous R-loop with the cloned mouse β -globin gene at a position corresponding to the two gene segments plus the intervening sequences. This was the first clear evidence that eucaryotic genes containing introns were functional. In his recent work, A. Schamböck (thesis, 1981) has converted 15 S β -globin precursor RNA into double-stranded cDNA, characterized it by restriction mapping and sequence analysis, and shown that it is an exact copy of the chromosomal DNA and contains the two intervening sequences. It has been convincingly shown that the 15 S precursor is the major component of globin premessenger RNA in various mammalian and avian erythroid cells. The existence of a larger precursor however is still controversial. Bastos and Aviv (1977) have described a 27 S precursor for mouse β -globin mRNA which has not been detected by other groups. Even under hybridization conditions at low stringency Haynes et al. (1978) were not able to hybridize RNA larger than 16 S to either α - or β -globin cDNA containing plasmids. In spite of these discrepancies, Scherrer et al. (1979) still believed the existence of a 27 S precursor and even postulated, based on their own data, the presence of a primary transcript

12 to 15 kb long. The R-loop experiments of Tilghman et al. (1978), indicating the colinearity between the 15 S precursor and the globin gene, suggests that a transcript larger than 15 S would have to extend further 3' or start further 5' than the 15 S precursor. Weaver and Weissmann (1979), in a modification of the S_1 mapping procedure of Berk and Sharp, have shown that the 5' end of the 15 S mouse β -globin pre-mRNA coincides with that of the mRNA. Recently, Grosveld et al. (1981) obtained a similar result in the case of the rabbit β -globin transcripts. In addition, in vitro synthesis from the β major globin gene also initiates at the cap site (Luse and Roeder, 1980), suggesting that site as the probable initiation site in vivo. Inspection of sequences preceding the cap site of eucaryotic messengers have revealed specific elements with conserved DNA sequences in homologous positions of many different genes, namely the TATA box, (Goldberg, thesis, 1979) located some 30 nucleotides upstream from the 5' end of the mature messenger, the CAAT box (Efstratiadis et al., 1980), some 80 nucleotides upstream from the cap site and a region around position -100 (Dierks, personal communication). The functional significance of these regions has been monitored by introducing into the elements sets of deletions or modifications. Dierks et al. (1981) for instance were able to show the importance of these putative promotor site components for both quantitative and qualitative regulation of transcription. These results make the hypothesis of a globin mRNA precursor starting 5' of the cap site very unlikely. Should a globin transcript larger than 15 S in fact exist, the most probable explanation would be a

continuation of RNA polymerase beyond the poly(A) addition site. Such cases have been reported for the major late transcript and the two early transcripts of Adenovirus as well as the late SV40 transcription unit (Nevins and Darnell, 1978; Ford and Hsu, 1978; Fraser et al., 1979; Nevins et al., 1980). Recently Hofer and Darnell (1981) were able to show that a majority of the RNA synthesis from the β major globin gene in induced erythroleukaemic cells begins at or near the cap site and that the majority of RNA polymerases transcribe at least a thousand bases past the poly(A) site. They also find a minority of RNA molecules which form an RNase resistant hybrid with a restriction fragment located immediately upstream of the cap site. They did not prove however, that this RNA is bound to the main body of the globin transcript, nor did they show to what extent its sequence is complementary to the sequence of the DNA. If this statement proves to be true, it would still represent only a rare event when compared to the amount of RNA starting at the cap site. Therefore, the current view on eucaryotic mRNA synthesis may be summarized as follows: The transcription starts at the cap site and RNA polymerase moves downstream, reading across the intervening sequences and stopping downstream of the poly(A) addition site. The primary transcript is capped either during or immediately after transcription at the 5' end, cleaved and polyadenylated at the 3' end and the transcripts of the introns are exised and the adjacent exon transcripts are ligated to the final protein coding body. After splicing, which is accomplished in the cell nucleus, mature messenger is transferred to the cytoplasm. Finally,

not every larger RNA molecule has to be a precursor or a splicing intermediate. Some of these molecules could represent erroneous transcription products which would never be used as active messenger but probably would be destroyed in the cell nucleus and thus show the same type of decay kinetics as genuine precursor molecules. In order to be able to call a high molecular weight RNA a precursor, it should for example, be purified and processed to the mature messenger in an in vitro splicing system. Several hypotheses on the biological significance of discontinuous genes and RNA splicing have been advanced. For Gilbert (1978) discontinuous genes might accelerate evolution by facilitating the formation of new proteins by combining parts of old ones. According to Tiemeier et al. (1978) intervening sequences might stabilize duplicated genes by acting as barriers to recombination. Darnell (1978) suggested that interrupted genes might represent a primitive genetic structure inherited from the common ancestor of procaryotes and eucaryotes. One important finding by Hamer and Leder (1979) was that SV40 recombinants carrying parts of the mouse globin gene were only able to produce stable RNA if they retained at least one functional intron, regardless of its viral or chromosomal origin. These data imply that splicing is a prerequisite for stable RNA formation. Supporting this theory is the finding of Gruss et al. (1979), who constructed an SV40 mutant that lacks the 16 S mRNA intron and were able to show that it fails to accumulate late mRNA. Such a mutant, however, could be "rescued" by the insertion of a heterologous splice site (Gruss and Khoury, 1980). One possible explanation for this

phenomenon is that splicing might be linked with the transport of RNA from the nucleus to the cytoplasm. In the case of histone messenger RNAs (Schaffner et al., 1978) or interferon messenger RNAs (Nagata et al., 1980) a splicing-independent transport mechanism would have to be postulated. Recently, however, Gruss et al. (1981) have shown that an "intronless" recombinant DNA molecule carrying the coding region and 3' flanking sequences of the rat prepro-insulin I gene replacing the late region of an SV40 viral vector was able to direct the synthesis of unspliced stable hybrid transcripts. It therefore seems that certain RNA structures can bypass the requirement for RNA splicing. A special phenomenon appears in DNA tumor viruses such as Polyoma and SV40, where the RNA splicing mechanism allows the virus to use its genetic information to a maximum extent, namely by splicing one single precursor into different messenger RNAs which will each be translated into a different protein. Similarly, Hagenbüchle and Schibler (1981) have been able to show that precursor RNA of the α -amylase gene was spliced differently in different organs, yielding messenger RNAs with different 5' non-coding regions (leader sequences). In the case of the immunoglobulins, split genes increase the possibility of generating antibody diversity (C-V joining; (Maki et al., 1980; Moore et al., 1981)) or allow the formation of different types of mRNA (for secreted and membrane bound antibodies) from one and the same precursor (Early et al., 1980). The mechanism of splicing is still unknown, however comparison of all the known intron sequences and their junctions have allowed to establish a consensus

sequence (Breathnach et al., 1978) which may be required for splicing. This "Chambon rule" states that the 5' border of an intron starts with a GT and the 3' border ends with an AG. These two nucleotides on either end of the intron however do not seem sufficient as signal for splicing, since the probability that they appear somewhere else in the coding sequences or are produced by a point mutation is too high to assure the stability of a gene. Comparison of the various α - and β -globin genes suggests that the sequences adjoining the splice sites have been most conserved during evolution. By contrast, there is almost no conservation of sequences in the introns. Therefore it would be conceivable that the coding nucleotides in the vicinity of the introns would be involved in secondary and tertiary structures in order to guarantee conjunction of the splice points. According to Breathnach et al. (1978) the requirements for such a model are met in only few of the known sequences. Another theory, recently proposed by several labs is based on an intermolecular RNA hybridization mechanism in which the necessary specificity and stereochemical orientation for splicing is provided by the hybridization of a separate RNA molecule to the precursor RNA. Several such "splicer molecules" have recently been proposed, involving either hybrid formations with the ends of introns (5' end of U_1 RNA: Lerner et al., 1980), exons (U_2 RNA or 4.5 S RNA: Ohshima et al., 1981) or involving both intron and exon sequences (VA_1 RNA: Murray and Holliday, 1979). It has been shown for most of these small nuclear RNAs that they are associated with proteins in form of ribonucleoprotein particles (RNPs). Lerner et al. have suggested

(1980) that the intact snRNPs could function as an RNase providing both the binding sites and the proteins effecting the catalysis, as in the case of RNase P (Stark et al., 1978).

1.3. The origin of this work

This work deals with a small RNA species present in mouse erythroleukaemic cells. As shown by Friend et al. (1971), mouse erythroleukaemic cells can be stimulated to differentiate along the erythroid pathway when exposed to DMSO at concentrations that do not inhibit cell growth. This finally results in the synthesis of high levels of hemoglobin. The possibility of turning on hemoglobin production made Friend cells an ideal system to study the globin gene transcription. When this work was originated, several labs had just shown the existence of a 15 S globin precursor RNA in various erythroid cells such as mouse fetal liver (Ross, 1976), anaemic spleen cells (Kwan et al., 1977) and erythroleukaemic Friend cells (Curtis and Weissmann, 1976; Bastos and Aviv, 1977). When Bastos and Aviv claimed that they were able to isolate a precursor larger in size than the one found in most laboratories, P. Curtis made several attempts to find globin-specific RNAs other than 15 S or 10 S in Friend erythroleukaemic cells. To this purpose, cells were pulse-labeled in vivo with inorganic phosphate, the total cell RNA was purified and separated according to size by sucrose gradient centrifugation. Each fraction of the gradient was analyzed separately for the presence of globin-specific sequences, using a recombinant plasmid containing the mouse chromosomal β -globin gene and surrounding regions as

hybridization probe. Instead of an RNA larger than 15 S, he identified a labeled RNA species of about 4 - 5 S which hybridized to the cloned DNA. Because of its size, it was suspected that it might be a piece of intervening sequence which had been spliced out from the 15 S precursor. To test this hypothesis, I analyzed the RNA by RNase T₁ fingerprinting, and compared it to the T₁ fingerprints obtained from the 15 S globin precursor by Curtis et al. (1977). This analysis showed that the small RNA species was not derived from the intervening sequences. However, it has the same T₁ fingerprint as RNA 16 found by Peters et al. (1977) in Moloney Murine Leukaemia virus infected 3T3 cells. By Southern hybridization techniques the small RNA was found to hybridize to a sequence starting 2300 bp downstream of the globin gene. A partial sequence analysis of the RNA was carried out using the conventional degradation procedures described in the Methods section. The DNA fragment hybridizing RNA 16 was characterized by restriction analysis and DNA sequencing. By comparison of the DNA sequence with the partial RNA sequence and the complete sequence of an RNA with identical fingerprint established by Harada and Kato (1980), it was shown that the two sequences were not identical but only partially homologous. In order to establish a model for the possible structure of the RNA-DNA hybrid, footprinting experiments were carried out. Hoping to get information about the possible role of the small RNA in the cell, we analyzed its association with several cell components such as nucleus, cytoplasm, poly(A) RNA and RNP particles.

2. MATERIALS AND METHODS

2.1. Materials

Medium for Friend cells was made according to Curtis and Weissmann (1976). Phosphate free SF-12 medium was prepared according to Ham (1965) by omitting sodium phosphate and using horse serum dialysed against 0.14 M NaCl, 50mM Tris HCl pH 7.5 .

Cells were grown in plastic petri dishes from Gibco Biocult. H3-Uridine and carrier-free (^{32}P) phosphate were purchased from New England Nuclear. When applied to the medium in large amounts, the (^{32}P) phosphate was first passed through Dowex 50- H^+ , adsorbed to Dowex I- Cl^- and eluted with 0.2 N HCl in presence of 10mM inorganic phosphate carrier.

Pancreatic RNase A was purchased from Worthington, RNase U_2 was from Sankyo; RNase T_1 (Calbiochem.) was treated with acid EDTA according to Dahlberg (1968). Partially purified RNase T_2 was prepared from Taka Diastase (Sankyo) as described by Hiramaru et al. (1966).

Pronase, 8 grade (Calbiochem.) was incubated at 20mg/ml in 5 mM EDTA, 10 mM Tris HCl pH 7.5 for 90 min at 37°C to destroy RNase activity.

An extracellular ribonuclease from Physarum polycephalum was a gift from F. Meyer. S_1 nuclease (11'000 U/ml) was a gift from A. Schamböck.

Electrophoretically pure DNase I was purchased from Worthington and treated with sodium iodoacetate by the procedure of Zimmermann & Sandeen (1966) to destroy residual RNase.

Restriction endonuclease BspI was prepared by A. Kiss, EcoRI by W. Boll, AluI, HinfI and PstI were purchased from Biolabs and TaqI from Porton.

Yeast carrier RNA from BDH Chemicals, (Poole, England) was made RNase free by extraction with phenol and precipitation with ethanol.

Bovine serum albumin (BSA) fraction V, B grade (Calbiochem.) was shaken with macaloid (Baroid National Lead Co., Houston, Texas) to remove RNase activity and dialysed for several days against distilled water.

(^{32}P) ATP was made according to I. Kennedy (unpublished) and was a gift from P. Dierks.

Guanosine tetraphosphate (pppGp) was made by carrying out an RNA polymerase reaction in presence of α -(^{32}P) GTP with sonicated calf thymus DNA as a template followed by digestion with nuclease T_2 .

^{125}I -labeled mouse globin mRNA was prepared according to Scherberg and Refetoff (1973).

Unlabeled mRNA was a gift from Dr. Williamson.

The mouse chromosomal β -globin plasmid DNA was prepared by an modification of the method of Currier and Nester (1976). It was originally cloned into the EK2 vector $\lambda\text{gt-WES}\lambda\text{B}$ by Tilghman et al. (1977) and later transferred into the EcoRI site of pBR322 by W. Boll of this laboratory. The plasmid, called Z-pBR322/Mchr βG -36 was grown initially in E.coli X 1776, later in E.coli HB101.

Homomix was made according to the procedure of Volckaert et al. (1976) and the yeast RNA used was from BDH Chemicals (Poole, England).

Cellulose acetate strips were from Schleicher and Schüll and PEI Cell 300 thin layer plates from Macherey-Nagel & Co.

DEAE - cellulose paper and 3MM sheets were from Whatman.

Nitrocellulose filters of 0.45 μ pore size (type HAW P 25) were from Millipore Co. and filters used for Southern transfers were purchased from Schleicher and Schüll.

Acrylamide (puriss.) and N'N methylene bisacrylamide were from Serva, Germany.

Phenol was distilled and stored at -20°C. SDS was recrystallized from aqueous ethanol prior to use.

Oligo T cellulose was purchased from P-L Biochemicals (Wisconsin).

2.2. Methods

2.2.1. Tissue Culture

Friend cells, clone 707/M2 were grown in plastic petri dishes in a 5% CO₂/air mixture at 37°C using supplemented Ham's F-12 medium (Curtis and Weissmann, 1976). When the cells were labeled for a short period (20 min or 30 min) with (³²P) inorganic phosphate, they were preincubated in phosphate-free SF-12/10% horse serum for 2 to 4 h. For labelling times of 18 and 2.5 h, the cells were suspended at a density of 2-3 x 10⁶ cells/ml,

pulse labelling experiments with (^{32}P) inorganic phosphate were carried out at a density of $5 \times 10^6 - 1 \times 10^7$ cells/ml. For induction of hemoglobin synthesis, cells were plated out at 2×10^5 cells/ml SF-12/10% horse serum containing 1.5% Me_2SO .

2.2.2. Total cell RNA isolation

Cells were pelleted from the medium in a (clinical) table centrifuge and taken up in cold 20 mM Tris HCl pH 7.5, 1 mM EDTA at a concentration of $2 - 5 \times 10^6$ cells/ml. SDS was added to a concentration of 2%, the suspension digested with 200 μg pronase/ml during 5 min at 37°C and passed through a syringe to shear the DNA clumps. The solution was extracted 3 times with 1 volume of phenol equilibrated with 20 mM Tris HCl pH 7.5, 1 mM EDTA at room temperature. After each extraction the top of the phenol phase was rinsed with 1/100 volume of Tris/EDTA buffer. The aqueous phase was adjusted to 0.3 M sodium-acetate pH 5.4 and the nucleic acids were precipitated with 2.2 volumes of ethanol. The pellet was redissolved in H_2O , adjusted to Tris HCl pH 7.5, 10 mM MgCl_2 , 2 mM CaCl_2 and incubated at 37°C for 10 min in presence of 5 - 15 μg DNase I treated with iodoacetate according to Zimmermann and Sandeen (1966). Excess EDTA was added, the solution was adjusted to 0.5% SDS, extracted once with phenol and purified on a G-100 Sephadex column (volume=3ml) with a small layer of Chelex-100 at the bottom. The excluded RNA peak was pooled, adjusted to 0.3 M sodium acetate pH 5.4 and precipitated with 2.2 volumes of ethanol.

2.2.3. RNA sequence determination

2.2.3.1. RNase T₁ and RNase A fingerprints

Samples of in vivo labeled (³²P) RNA 16 were digested with RNase T₁ (0.5 U/μg RNA in 20 mM Tris HCl pH 7.5, 2 mM EDTA) or with RNase A (1μg/20μg RNA in 20 mM Tris HCl pH 7.5, 2 mM EDTA), and the resulting oligonucleotides were separated by 2-dimensional electrophoresis according to Sanger et al. (1965). For determination of the sequence, the spots were extracted by 2 M Triethylammonium carbonate pH 10 and subjected to a second digestion with another enzyme.

2.2.3.2. Sequence determination of the RNase T₁ and RNase A oligonucleotides

The nucleotide sequence of the oligonucleotides from the fingerprints were determined according to Brownlee (1972). The T₁ oligonucleotides were further degraded with:

- a) RNase A (1μg/μg RNA) in 10μl 20 mM Tris HCl pH 7.5 at a final concentration of 10μg RNA/10μl for 1 hour at 37°C.
- b) RNase U₂ (1 U/10μg RNA) in 10μl 50 mM NaAc pH 4.5, 2 mM EDTA, 0.1 mg/ml BSA at a final concentration of 10μg RNA/10μl for 2 hours at 37°C.
- c) Physarum enzyme (3.75 U/10μg RNA) in 10μl 0.1 M NaAc pH 5.0, 1 mM EDTA, (50μg/ml) BSA at a final concentration of 10μg RNA/10μl for 30 min and 2 hours at 37°C.
- d) RNase T₂ (50 U/ml) (2 volumes) plus RNase A (5 mg/ml) (1 volume) plus 15 mM NaAc pH 4.5, 1 mM EDTA, (1 mg/ml) carrier RNA (21 volumes) at a final concentration of 10μg RNA/10μl for 1 hour at 37°C.

- e) Alkaline Phosphatase (0.02 U/10 μ g) RNA in 10 μ l 50 mM Tris HCl pH 8.0, 1 mM MgCl₂ at a final concentration of 10 μ g RNA/10 μ l for 1 hour at 37°C.
- f) S₁ nuclease (11 U/0 - 10 μ g) RNA in 10 mM NaAc pH 4.5, 1 mM ZnCl₂ at a final concentration of 10 μ g RNA/10 μ l for 30 min at 37°C.

Oligonucleotides obtained by RNase A digestion were further degraded with:

- a) RNase T₁ (1 U/ μ g RNA) in 10 μ l 20 mM Tris HCl pH 7.5, 1 mM EDTA at a final concentration of 10 μ g RNA/10 μ l for 30 min at 37°C.
- b) RNase T₂ as described above.

The mononucleotides obtained by digestion with RNase T₂, were separated by electrophoresis on Whatman 3MM paper at pH 3.5. RNase U₂ and Physarum RNase digests were fractionated by electrophoresis on DEAE paper at pH 1.9. RNase T₁ and RNase A double digests were analyzed by electrophoresis on DEAE paper at pH 3.5. After digestion with nuclease S₁, the products were analyzed by thin layer chromatography on PEI cellulose by two dimensional separation according to Randerath and Randerath (1964). Alkaline phosphatase products were separated by electrophoresis on Whatman 3MM paper at pH 3.5.

2.2.4. Restriction analysis, Southern hybridization and DNA sequence determination

The restriction map of Z-pBR322/Mchr β G-36 was established by complete digestion with restriction endonucleases BspI, AluI,

HinfI and TaqI and by mapping experiments according to Smith and Birnstiel (1976). Transfer of restricted DNA to Millipore filters and hybridization with labeled probes were carried out according to the method of Southern (1975). DNA sequencing was essentially done as described by Maxam and Gilbert (1977), except that the DNA was not denatured before the kinase reaction. Prior to sequencing the labeled fragments were either strand separated on a PAA gel or cleaved with a second restriction enzyme and the resulting fragments separated by gel electrophoresis. Agarose gels were stained with ethidium bromide and photographed under UV through a red filter with a Polaroid camera. Autoradiography was usually carried out by using a tungstate intensifier screen and exposing the filter at -70°C .

3. RESULTS

3.1. Discovery of RNA 16 in Friend cells

As mentioned above, P. Curtis identified an RNA species in the 4 - 5 S gradient fractions of Friend cell RNA hybridizing to Z-pBR322/Mchr8G-36 plasmid DNA. In order to characterize the small RNA more closely, a number of short- and long-term labelling experiments were carried out.

Cells were labeled with (^{32}P) phosphate for 30 min or 16 - 18 h. Total cell RNA was heat denatured and centrifuged through a 5% - 23% sucrose gradient (Fig. 1). The fractions corresponding to the 4 - 5S RNA peak were hybridized to a filter containing 18-20 μg Z-pBR322/Mchr8G-36 plasmid DNA. The hybridization was followed by 3 washes of 10 min in 1 ml 0.5 x SSC, 0.5% SDS at room temperature and a 2 h wash in 1 ml hybridization buffer at 45 $^{\circ}\text{C}$ in a shaking waterbath. The product was eluted from the filter and usually purified further by PAA gel electrophoresis. The material bound to the chromosomal globin DNA consisted of an apparently homogeneous main band and a few fainter species, larger in size (Fig. 2).

A method for the separation of RNA species by 2-dimensional PAA gel electrophoresis was described by Fradin et al. (1975). This system allowed a separation of RNA 16 into at least 4 spots, later shown to be due to a heterogeneity at the 3' end of the molecule (Fig. 3a and b). The samples to be separated were either crude 4 - 5S fractions obtained by sucrose gradient sedimentation of total heat-denatured Friend cell RNA or were obtained by further purification of this RNA by hybridization to a Millipore filter containing Z-pBR322/Mchr8G-36 plasmid DNA.

Figure 1 Sucrose gradient centrifugation of labeled Friend cell RNA.

Total (^{32}P) RNA from uninduced Friend cells labeled for 30 min (●-●-●) and induced Friend cells labeled for 16 h (x-x-x) was denatured for 45 sec at 100°C and centrifuged through a 5% - 23% sucrose gradient in 50 mM Tris HCl (pH 7.5), 5 mM EDTA, 0.1% SDS for 16 h at 15°C at 35'000 revs/min in an SW41 Spinco rotor.

Figure 2 PAA gel electrophoresis of (^{32}P) RNA from induced Friend cells.

RNA < 10 S from cells labeled for 30 min or 16 h was hybridized to filters containing Z-pBR322/Mchr β G-36 DNA. After elution, the RNA was heated in 7M urea, 90 mM Tris-borate pH 8.2 , 2.5 mM EDTA at 70°C for 2 min before electrophoresis through a 6% polyacrylamide gel as described by Maniatis et al. (1975). In addition, the gel was overlaid with a 3% spacer gel containing 7M urea, 90 mM Tris-borate pH 8.2 , 2 mM EDTA. It was run in Tris-borate/EDTA buffer at 300 V for 4 h and subjected to autoradiography.

- a) β -globin mRNA applied as a size marker
- b) RNA from cells labeled for 16 h after elution from the Z-pBR322/Mchr β G-36 DNA
- c) RNA from cells labeled for 30 min after elution from the Z-pBR322/Mchr β G-36 DNA

Fig. 1

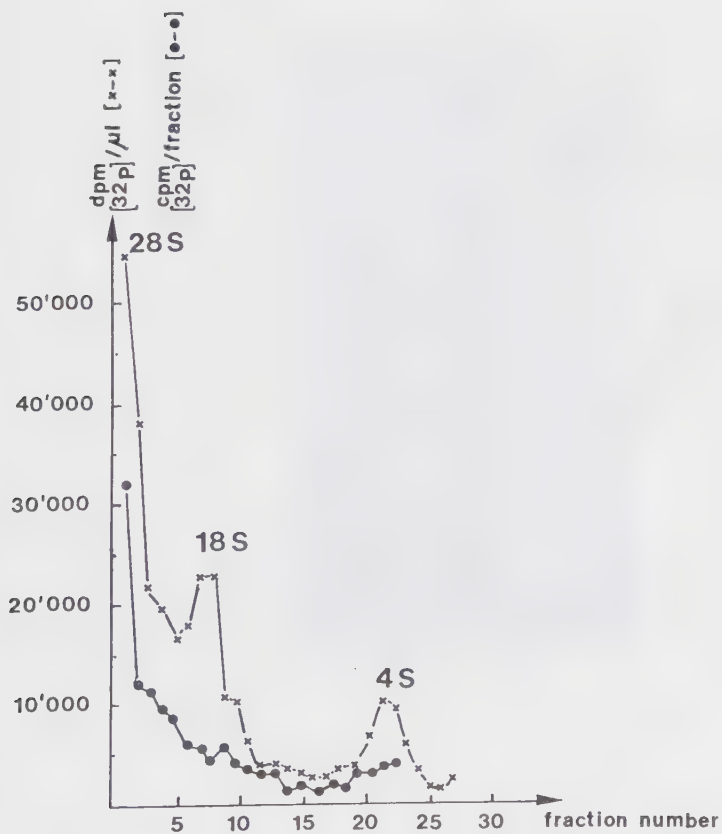


Fig. 2

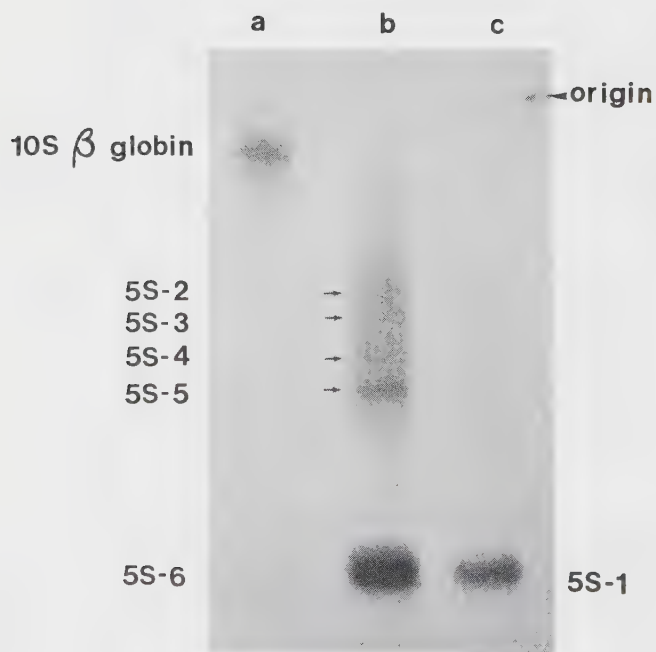
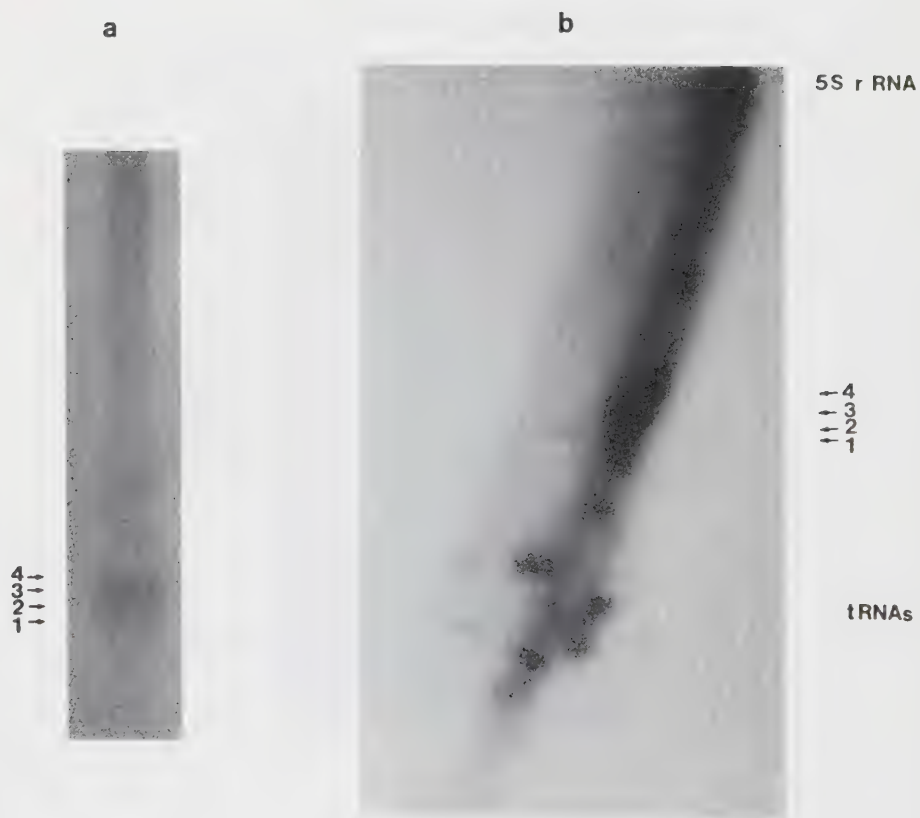


Figure 3 2-dimensional PAA gel electrophoresis of RNA 16 and small cellular RNAs using the method by Fradin et al. (1975).

- a) RNA 16 from Friend cells was purified by hybridization on a nitrocellulose filter containing Z-pBR322/Mchr8G-36 plasmid DNA and separated by PAA gel electrophoresis on the first dimension (Fradin et al., 1975); 1 through 4 correspond to RNA 16 with increasing numbers of U-residues at the 3' end of the molecule.
- b) 2-dimensional gel electrophoresis (Fradin et al., 1975) of low molecular weight Friend cell RNA eluted from poly(A) RNA > 18 S. For experimental details on RNA preparation see legend to Fig. 10. The position of RNA 16 is shown by arrows.

Fig. 3



3.2. Yields of (^{32}P) RNA 16 from Friend cells under various labelling conditions

In order to compare the labelling rate of RNA 16 in relation to the bulk of the cellular RNA, cells were incubated for 20 - 30 min, 2.5 h or 16 - 18 h with (^{32}P) inorganic phosphate and the RNA was extracted, heat denatured and fractionated by sucrose density gradient centrifugation as indicated in the Methods section. The 4S - 5S gradient fractions were pooled and hybridized to Z-pBR322/Mchr β G-36 bound to nitrocellulose filters. The bound fraction was eluted from the filter and separated by electrophoresis on a denaturing PAA gel. In order to ascertain that the amount of DNA on the filter was not a limiting factor, the RNA was eluted from the filter after a first hybridization and the bulk of the cellular RNA was rehybridized a second time to the filter. However no significant amount of RNA was bound. Nevertheless, the yields of RNA 16 from several experiments shown in table 1 allow only a rough estimate of the quantities present, since the experiments were not carried out in parallel. Any decrease of the yield of RNA due to a loss of DNA from the filter was neglected. Another reason why the estimates of RNA 16 in table 1 might be too low may be the following: As will be discussed later, RNA 16 is bound to poly(A)-containing high molecular weight RNA (≥ 18 S) from which it can be released by melting. During these first sets of experiments, no step was taken to ensure complete heat denaturation of the samples before sucrose gradient sedimentation, so that at least a portion of RNA 16 could have remained bound to the high molecular weight fraction.

Table 1 Amounts of labeled small RNA (RNA 16 + related species) using different labelling times.

Induced Friend cells were labeled with (^{32}P) inorganic phosphate in phosphate free medium as indicated in the Methods section. (^{32}P) total cell RNA was isolated, heat denatured for 45 sec and sedimented through a 5% - 23% sucrose gradient. The 4 - 5S RNA fractions were pooled and hybridized as indicated under 3.1. to a Millipore filter containing 18 μg - 20 μg Z-pBR322/Mchr β G-36 plasmid DNA. The hybridized material was further purified on a denaturing PAA gel (described in the legend to Fig. 2). The values were corrected for (^{32}P)-decay, 10^7 cells contained about 100 μg total cell RNA. In long-term labelling experiments, the specific activity of total cell RNA was about 1×10^6 dpm/ μg ; in pulse labelling experiments about 0.2×10^6 dpm/ μg .

Table 1

time of labelling	^{32}P (mCi)	N ^o of cells	tot. RNA (dpm)	RNA hybri- dized to Z-pBR322/ Mchr8G-36	RNA 16
				tot. RNA (%)	tot. RNA (%)
30 min	100	9×10^7	2.2×10^9	0.029	0.0053
20 min	100	9×10^7	0.4×10^9	0.018	0.0056
2.5 h	20	9×10^7	2.4×10^9	0.038	0.0156
16 h	15	9×10^7	3.8×10^9	0.018	0.0010
18 h	18	1.2×10^8	6.0×10^9	0.026	0.0046

3.3. Comparison of the various small RNA species by fingerprint analysis

To further analyze this set of RNAs and to find out whether they were related, they were looked at by fingerprinting. If they were composed of intron sequences we would expect to find some of the larger oligonucleotides specific for the 15 S globin precursor, but absent in the 10 S mRNA (Curtis et al., 1977).

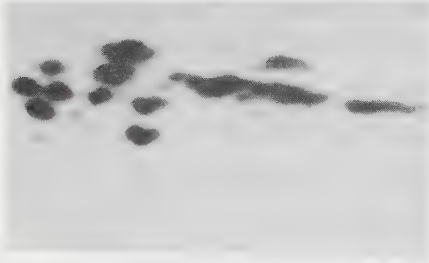
Each RNA was digested with RNase T₁ and the oligonucleotides were fractionated by cellulose acetate electrophoresis at pH 3.5 followed by homochromatography on PEI plates (Volkaert et al., 1976). The patterns were compared to one another as well as to the intron-specific spots of Curtis et al. (1977). The results are shown in figure 4 a-i. The smallest species, 5S-6, which was also the most prevalent one and corresponded to RNA 16, was further analyzed. 5S-3 and 5S-4 seem to have identical fingerprints, 5S-5 has three (24, 25, 26) and 5S-2 has three other (5, 17, 20) additional spots as compared to 5S-3 and 5S-4. There was no obvious similarity to the intron-related oligonucleotide pattern (i.e., to the oligonucleotides occurring in 15 S but not in 10 S RNA) of Curtis et al. (1977). The fainter set of bands did not seem to be related to the main species, since their large oligonucleotides were different. We obtained evidence from Southern hybridization experiments that 5S-4 binds to the same DNA region on the mouse chromosomal globin plasmid (see section 3.6.2.) as 5S - 6. Some faint spots visible in the fingerprints are probably contaminants due to the background

Figure 4a - i PEI fingerprint analysis of the small RNA species binding to Z-pBR322/Mchr8G-36 DNA.

(^{32}P) RNA binding to Z-pBR322/Mchr8G-36 was separated on a 6% PAA gel with 3% spacer gel (legend to Fig. 2). (^{32}P) RNA corresponding to the various gel bands was eluted in 1 ml 1 x TNE at 37°C for 2 times 1 h in presence of yeast carrier RNA and digested with 3 μl RNase T_1 (1 unit/ μl). Oligonucleotides were separated by homochromatography according to the procedure of Volckaert et al. (1976). a) fingerprint of the main species in the pulse-labelling experiment, 5S-1, b) main species of the long-term labelling experiment, 5S-6, c) tracing of the main species, d) 5S-2, e) 5S-3, f) 5S-4, g) 5S-5, n) composite tracing of the four large species, i) composite tracing of the 10 S globin mRNA and the 15 S precursor as published by Curtis et al. (1977). The terminology for the different RNA species is the same as in Fig. 2.

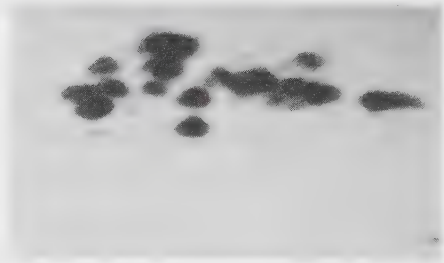
Fig. 4a-i

a



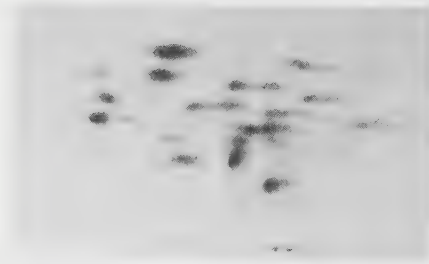
5S-1

b



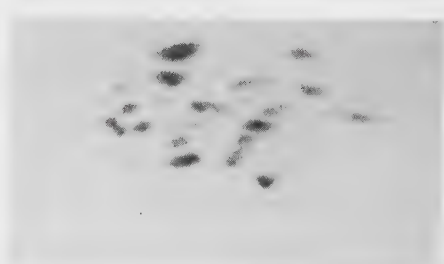
5S-6

d



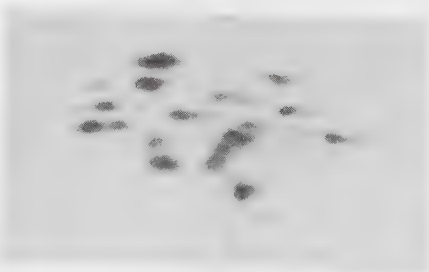
5S-2

e



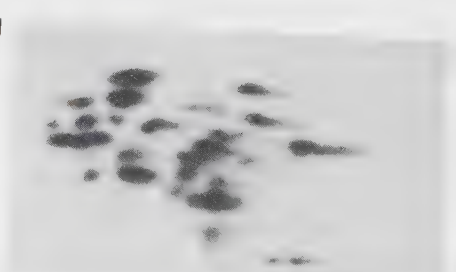
5S-3

f



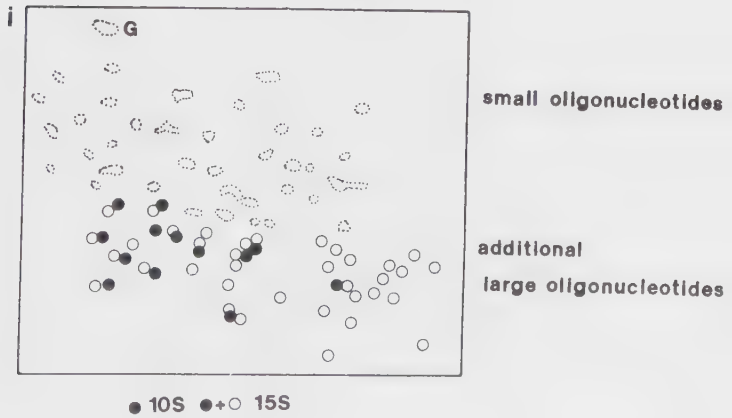
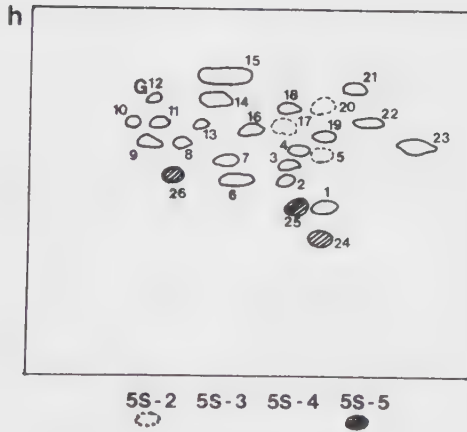
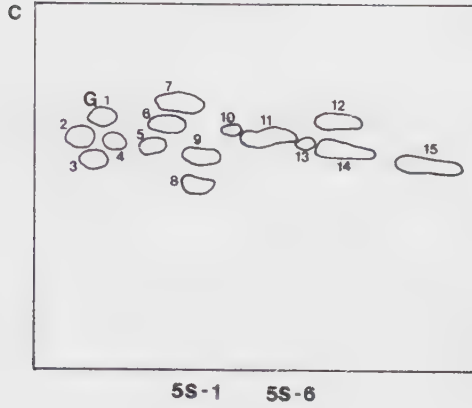
5S-4

g



5S-5

Fig. 4a-i (cont.)



smear on the gel. As expected 5S-1 of the pulse labeled material and 5S-6 of the long-term labeled material have an identical fingerprint.

3.4. Partial sequence of RNA 16

In view of the fact that the small RNA species described above were not related to intron sequences, a partial analysis of their sequence was carried out. Of particular interest were the possible modifications at the 5' end and 3' end of the molecule, which might allow some conclusions about its role in the cell.

3.4.1. RNase T₁ and RNase A fingerprints

Since the ratio of (³²P) labeled RNA 16 to total cell RNA seemed more favorable after 2.5 hours of labelling rather than overnight, RNase T₁ fingerprints were made with RNA from short labelling times.

Samples of purified uniformly labeled RNA 16 were digested with RNase T₁ and RNase A and the resulting oligonucleotides were separated by 2-dimensional electrophoresis according to Sanger et al. (1965). This method of separation was used rather than chromatography on PEI cellulose plates, in order to facilitate a comparison with published data on other RNAs. Autoradiographs of RNase T₁ and RNase A fingerprints of RNA 16 are shown in Fig. 5 a-d. The various digestion products were further analyzed to establish their sequences as described in the Methods section.

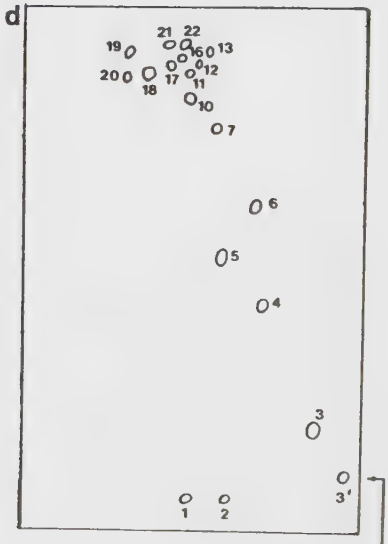
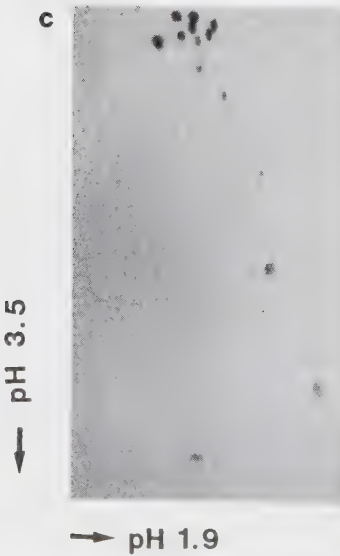
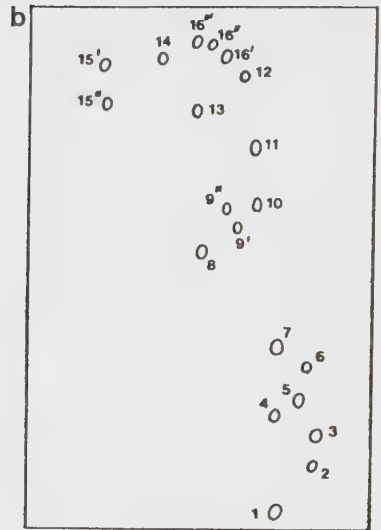
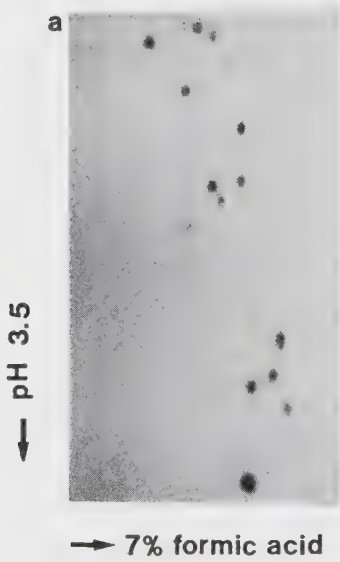
3.4.2. Catalogues of the RNase T₁ and RNase A oligonucleotides

The RNase T₁ and RNase A oligonucleotide catalogues are shown in tables 2 and 3, a detailed derivation of the sequences

Figure 5a - d RNase T_1 and RNase A 2-dimensional fingerprints of RNA 16 according to the method of Sanger et al. (1965).

a) RNase T_1 fingerprint of RNA 16, b) tracing of (a), c) RNase A fingerprint of RNA 16 (the spot corresponding to C does not show on the fingerprint. It could have been cut off during the manipulation), d) tracing of (c).

Fig. 5a-d



cut out from the
1st dimension or
bad transfer

Table 2 RNase T₁ oligonucleotide catalogue of RNA 16.

The details on the enzymatic cleavages of each oligonucleotide are shown in the Appendix.

Table 3 RNase A oligonucleotide catalogue of RNA 16.

The details on the enzymatic cleavages of each oligonucleotide are shown in the Appendix.

Table 2 RNase T₁ oligonucleotide catalogue of RNA 16

	sequence our data	Harada and Kato (1980)	<u>molar yields</u>		
			our data	Harada and Kato (1980)	theo- retical
1	G	G	13.4	13.1	13
2	CG	CG	0.8	0.8	1
3	C ₂ G	C ₂ G	1.8	1.6	2
4	AG	AG	3.7	4.6	5
5	CAG	CAG	2.0	1.8	2
6 ^{a/b}	CACG CCAG	CACG CCAG	1.8	1.7	2
7	A ₂ G	A ₂ G	1.0	1.0	1
8	UG	UG	2.3	2.4	2
9'	CUG	CUG	1.3	1.0	1
9''	UAG	UAG	2.3	2.4	2
10	C ₂ UG	C ₂ UG	1.4	0.9	1
11	AUCACG	AUCACG	1.0	1.0	1
* 12	ACU ^a U ₂₋₅ N ^b OH	CUACACA ^c U ₂₋₅ U ^d OH	1.9	1.1	1
* 16'	or				
* 16''	CUACACA ^c U ₂₋₅ N ^b OH				
* 16'''					
13	U ₂ CG	U ₂ CG	1.5	1.2	1
14	AU ₃ G	AU ₃ G	1.5	1.2	1
** 15'	pppGp	pppGp			
** 15''	ppGp	ppGp	0.9	0.9	1
		p3p			

* Spots 12 and 16' to 16''' are different oligonucleotides derived from the heterogeneous 3' end.

** 15' and 15'' are derived from the 5' end.

Table 3 RNase A oligonucleotide catalogue of RNA 16

No.	our data	Harada and Kato (1980)	Molar yield		
			our data	Harada and Kato (1980)	theoretical
1	U	U	4.2	7.4	7-10
2	U cyclic		3.2		
3	AC	AC	2.8	3.8	4
3'	C	C	2.9	6.6	6
4	GC	GC	2.7	3.0	3
5	AU	AU	0.9	1.3	1
6	AGC	AGC	0.8	1.0	1
7	G ₂ C	G ₂ C	1.3	1.0	1
10	AGU	AGU	1.2	1.1	1
11	G ₃ C	G ₃ C	1.0	0.8	1
12	(AG)(G) ₂ C	GAGGC	0.9	1.0	1
13	(AG) ₂ GC	AGAGGC	1.0	1.0	1
16	(AG)(G)AU	AGGAU	1.1	1.2	1
17	(AG)(G)U	GAGU	1.0	1.2	1
18	G ₂ U	G ₂ U	4.0	3.4	1
		pGpC			
19	ppGpC	ppGpC	0.7	1.1	1
20	pppGpC	pppGpC			
21	(AG) ₂ (G) ₂ AU	AGAGGGAU	1.3	1.2	1
22	(A ₂ G)(AG) ₁₋₂ (G) ₂₋₃ C	GAAGGAGGC	0.8	1.0	1

of each oligonucleotide is given in the Appendix. The estimation of the molar yields of oligonucleotides may be subject to some error due to the unequal specific activity of the nucleotide pools during the relatively short labelling periods (2.5 hours). As shown in table 2 and 3, RNA 16 has a very high content of G residues (39.5% - 38.3%). By S_1 digestion followed by 2-dimensional chromatography on thin layer (Methods section) the 5' terminal nucleotide could be identified as pppG., indicating that the RNA might be a primary transcript. The 3' terminal oligonucleotide (i.e. the only RNase T_1 product not ending with a G residue) appeared on the fingerprint as a series of at least 4 spots, the only difference being the number of U residues at the 3' end of the molecule (2 to 5 labeled Up). The 3' terminal nucleoside could not be determined with the methods used here (see Appendix). Since the molecule was very G rich, it was not possible to match the RNase T_1 and RNase A data to determine the order of the oligonucleotides.

3.4.3. Comparison of the data with the complete sequence of RNA 16 published by Harada and Kato (1980).

Recently Harada and Kato (1980) published the complete sequence of RNA 16 (Fig. 6), which had been isolated from mouse kidney cells (C3-H2-K), mouse lymphoma cells (L 1210/C) and baby hamster kidney cells (BHK-21). The data of tables 2 and 3 agree with the published sequence. As shown in table 2, the structure of the 3' T_1 oligonucleotide could not be determined unambiguously by further enzyme digestions. The correct structure is shown in the sequence determination by Harada. As pointed out by the authors, there are a few interesting features about

Fig. 6



Figure 6 Sequence of RNA 16 as published by Harada and Kato (1980). Underlined are some repeats and the sequence typical for the Alu family of interspersed sequences.

the sequence of RNA 16. It contains 2 sets of repeated sequences (GCCGGUAG : residues 1 to 8 and 20 to 27) and GAGGCAGAGG : residues 40 to 49 and 46 to 55, overlapping). Another point of interest is the purine rich sequence in the middle of the molecule (20 out of 22 residues between nucleotide 36 and 57). This sequence is homologous to the Alu family of interspersed repeated sequences found in the mammalian genome (Houck, et al. (1979).

3.5. Occurrence and intracellular distribution of RNA 16

Small RNA molecules have been postulated to be active in controlling gene expression or RNA splicing. Information about the distribution of RNA 16 in the cell or its occurrence in different cell types could help to understand its biological function.

3.5.1. Correlation of the appearance of RNA 16 with cell induction

Induction of Friend cells with DMSO causes not only an activation of the globin gene expression but also several other changes in the cell metabolism (Friend et al., 1971). In order to determine whether the presence of RNA 16 was a result of DMSO induction, induced and non-induced Friend cells were labeled in a set of parallel experiments with (^{32}P) inorganic phosphate for 18 h or 30 min. The amounts of RNA 16 were measured as indicated in Fig. 7 and table 4. In the case of the long-term labelling experiment the relative amount of RNA 16 is similar for induced and non-induced cells. In pulse-labelling experiments there is however a four-fold increase of the re-

Figure 7 PAA gel electrophoresis of small RNA from induced and non-induced Friend cells labeled for different lengths of time.

2×10^7 uninduced and 1.5×10^7 induced Friend cells were labeled for 30 min with 20 mCi (^{32}P) inorganic phosphate. The total cell RNA was extracted, heat denatured and fractionated by sucrose gradient centrifugation (Methods section). The yield of RNA $\leq 18 \text{ S}$ was 1.3×10^7 dpm for the uninduced and 2.0×10^7 dpm for the induced cells (approximately 50% of the total cell RNA). 2.3×10^7 uninduced and 2.2×10^7 induced cells were labeled for 18 h with 4.5 mCi. The yields of RNA $\leq 18 \text{ S}$ was 1.5×10^9 dpm for the uninduced and 1.2×10^9 dpm for induced cells (36 - 40% of the total cell RNA). For each RNA preparation, various amounts of RNA were hybridized to Millipore filters containing 6.25 μg Z-pBR322/Mchr β G-36 DNA (to ascertain that the amount of DNA on the filter does not limit the extent of hybridization) as described in the Methods section. The RNA was then eluted from the filter and separated by electrophoresis on 6% PAA gel as shown in Fig. 2. Slots a) - c) of Fig. 7 represent increasing amounts of RNA from uninduced cells (18 h labelling), d) - f) induced cells labeled for 18 h, g) + h) uninduced cells labeled for 30 min, i) + k) induced cells labeled for 30 min.

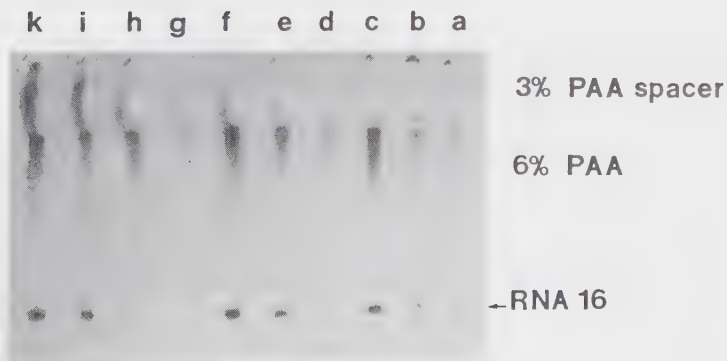
Table 4 Quantitative determination of RNA 16 and related species from induced and non-induced Friend cells.

Experimental procedure as in legend to Fig. 7. The input values of each hybridization experiment and the yields of RNA 16 are given in table 4. All values were corrected for radioactive decay of (^{32}P).

Table 4

sample	input into hybridization (RNA < 18 S) dpm	$\frac{(^{32}\text{P RNA 16})}{(^{32}\text{P RNA} \leq 18 \text{ S})}$ in gel
a)	528'000	0.045%
b)	1'055'000	0.050%
c)	1'981'000	0.036%
d)	503'000	0.069%
e)	1'000'000	0.046%
f)	1'992'000	0.039%
g)	584'000	0.019%
h)	1'115'000	0.016%
i)	606'000	0.081%
k)	1'191'000	0.072%

Fig. 7



lative amount of RNA 16 in induced cells over the uninduced ones. In addition the relative amount is much lower after a long period of labelling than after 30 min. This result can be interpreted in two ways: a) The production of RNA 16 is enhanced due to a less than average sensitivity of RNA 16 to the general inhibition effect of DMSO, b) The production of RNA 16 is enhanced by DMSO, however the product has a high rate of turnover, such that the stimulation is observed only at short labelling times. With increasing amounts of input RNA the relative amount of RNA 16 decreases slightly in every set. Possibly the amount of DNA on the filter is not in sufficient excess over the RNA hybridized.

3.5.2. Intracellular location of RNA 16

RNA involved in protein synthesis is expected to be found in the cytoplasm, in form of mRNPs or polysomes associated with membranes or in the soluble fraction of the cytoplasm. Some small RNA species appearing in the nucleus were believed to be associated with chromatin and possibly to be involved in gene regulation (Bonner and Widholm, 1967). Several other laboratories (Heyden and Zachau, 1971; Artman and Roth, 1971; Szeszak and Pihl, 1972) however failed to confirm the findings of Bonner et al., (1967) and today it is generally accepted that this chromatin-associated RNA is an experimental artefact. Another group of small RNAs was shown to migrate back and forth between the nucleus and the cytoplasm and was recently postulated to be involved in splicing (Lerner et al., 1980) and possibly in the transport of the mature mRNA from the nucleus to the cytoplasm (Harada and Kato (1980). In order to determine the distribution of RNA 16 between nucleus and cytoplasm, cells were labeled for 2.5 h with (^{32}P) inorganic phosphate and were fractionated as described by Penman (1966). As shown in Fig. 8

Figure 8 PAA gel electrophoresis of the nuclear and cytoplasmic small RNA fraction from induced and non-induced Friend cells.

3×10^7 induced or non-induced Friend cells were labeled for 2.5 hours with 3.9 mCi (^{32}P) inorganic phosphate and the nuclear and cytoplasmic fractions were separated in reticulocyte swelling buffer/NP40 as described by Penman (1966). The RNA was extracted and fractionated by sucrose gradient centrifugation as indicated in the Methods section. The yields of RNA $\leq 18 \text{ S}$ were 4.8×10^7 dpm for the nuclear fraction of non-induced cells, 9.5×10^7 dpm for the cytoplasmic fraction of non-induced cells, 7.7×10^7 dpm for the nuclear fraction of induced cells and 4.8×10^7 dpm for the cytoplasmic fraction of induced cells. As in experiment 3.5.1., increasing amounts of RNA of each fraction were hybridized to a filter containing 6.25 μg of 2-pBR322/Mchr β G-36 DNA, and the RNA hybridized to the filter was eluted and analyzed by electrophoresis on a 6% PAA gel. a) - c) represents increasing amounts of nuclear RNA from uninduced cells, d) - f) cytoplasmic RNA from uninduced cells, g) - i) nuclear RNA from induced cells, j) - l) cytoplasmic RNA from induced cells.

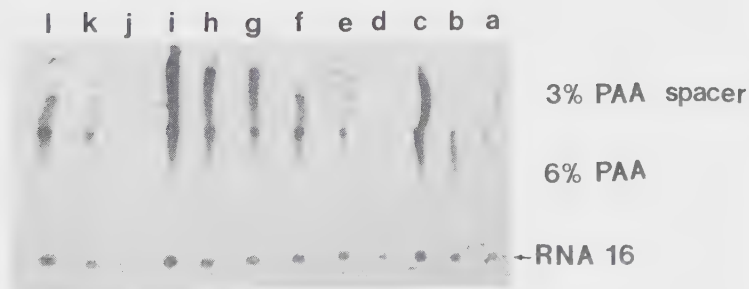
Table 5 Quantitative determination of RNA 16 in nuclei and cytoplasm of induced and non-induced Friend cells.

Experimental procedure as in Legend to Fig. 8. The input values of each hybridization experiment and the yields of RNA 16 are given in table 5. All values were corrected for radioactive decay of (^{32}P).

Table 5

sample	input into hybridization (RNA < 18 S) dpm	$\frac{(^{32}\text{P RNA 16})}{(^{32}\text{P RNA} < 18 \text{ S})}$ in gel
a) uninduced nuclear	585'000	0.102%
b) uninduced nuclear	1'254'000	0.081%
c) uninduced nuclear	2'486'000	0.061%
d) uninduced cytoplasm.	869'000	0.110%
e) uninduced dytoplasm.	1'355'000	0.075%
f) uninduced cytoplasm.	2'464'000	0.074%
g) induced nuclear	649'000	0.065%
h) induced nuclear	1'225'000	0.064%
i) induced nuclear	2'354'000	0.048%
j) induced cytoplasm.	638'000	0.085%
k) induced cytoplasm.	1'260'000	0.046%
l) induced cytoplasm.	2'523'000	0.061%

Fig. 8



and table 5, RNA 16 is clearly present in both cellular fractions in about equal quantities. This result, like those of most cell fractionation experiments, should be interpreted with some caution because the cellular fractions are usually cross-contaminated to various degrees. In particular the nuclei prepared with the technique used here are quite porous (Kato and Lowry, 1973). Small RNA molecules of the size of RNA 16 could be susceptible to leaking out of the nucleus. A confirmation of our result was obtained by Harada and Ikawa (1979), who found that RNA 16 was bound to both nuclear and cytoplasmic poly(A)-containing RNA.

3.5.3. Presence of RNA 16 in mouse L-cells

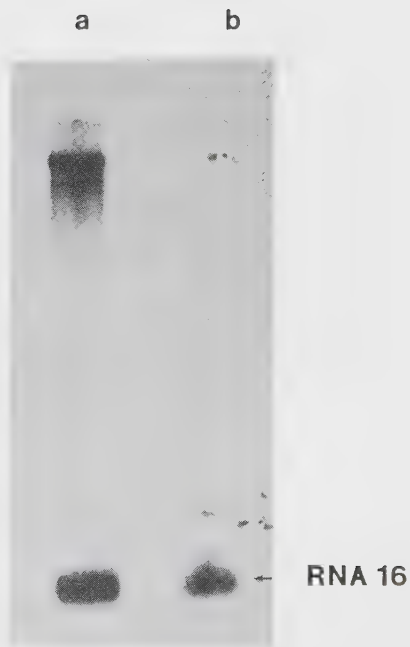
Peters et al. (1977) found RNA 16 associated to the 70 S complex of Moloney-Murine Leukaemia Virus RNA in a 3T3 cell producer cell line. Since Friend cells were originally also transformed with RNA tumor viruses (Friend et al., 1965), it seemed possible that the presence of RNA 16 was related to virus infection. Therefore, mouse L-cells, an untransformed fibroblast cell line, were analyzed for the presence of RNA 16. Friend cells and L-cells were labeled in parallel, their RNAs were hybridized to Z-pBR322/Mchr β G-36 and the bound material was electrophoresed side by side on the same PAA gel. As shown in Fig. 9, L-cells contain an RNA species which hybridizes to the β -globin DNA fragment and has the same mobility on gel electrophoresis as RNA 16. Later work by Harada et al. (1979) supported this result, since they found RNA 16 in infected and non-infected hamster, rat and mouse cells. It seems to be specific for rodent cells, since it is absent from cat, chicken and human cells.

Figure 9 PAA gel electrophoresis of small RNA from Friend and mouse L-cells.

Confluent mouse L-cells (a 10 cm $\varnothing$ plate) were labeled with 5 mCi (^3H), uridine overnight and 6×10^7 Friend cells were labeled for 2.5 h with the same amount of radioactive precursor. Friend cells incorporated 2.6×10^9 dpm into total cell RNA, the incorporation for L-cells was 7.6×10^8 dpm. The material was heat denatured, centrifuged through a 5% - 23% sucrose gradient, and RNA ≤ 18 S was hybridized to Z-pBR322/Mchr β G-36 DNA on Millipore filter. The bound material was electrophoresed on a 6% PAA gel (as legend to Fig. 2). After the electrophoresis the gel was subjected to fluorography in presence of PPO and DMSO according to Bonner and Laskey (1974).

a) Friend cell small RNA, b) L-cell small RNA.

Fig. 9



3.5.4. Association of RNA 16 with high molecular weight poly(A)

RNA

Following a suggestion by R. Kamen, we tried to isolate RNA 16 by thermal release from high molecular weight poly(A)-containing RNA from Friend cells. Total cell RNA was applied on a sucrose gradient without previous heat denaturation. The gradient was divided into three fractions (RNA < 18 S, RNA between 18 S and 28 S, RNA $\geq$ 28 S). These fractions were independently passed through oligo(dT) cellulose. The poly(A)-containing fraction of each group of RNA was heat denatured and passed through a second oligo(dT) column. The flowthrough fraction of the high molecular weight RNA was shown by electrophoresis on a PAA gel to contain RNA 16. The result of the experiment is shown in Fig. 10. Because of the strong exposure of the fourth lane it is not possible to decide whether the typical pattern corresponding to RNA 16 (Fig. 3a) is also present in the low molecular weight RNA. Even though these findings speak for an interaction of RNA 16 with poly(A) RNA, the possibility of an experimental artefact due to aggregation of RNA during phenol extraction cannot be excluded.

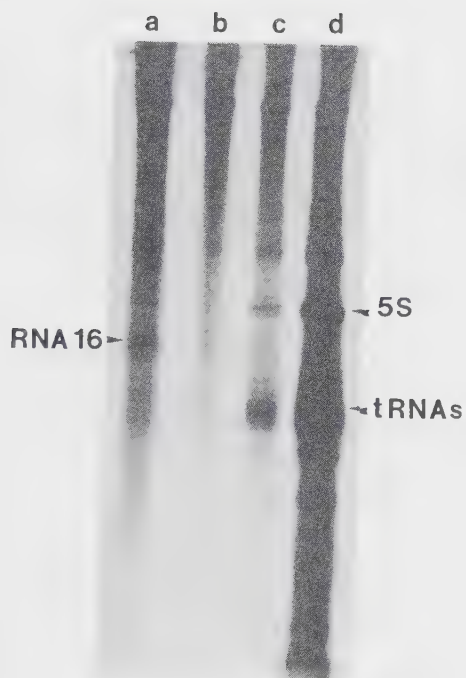
3.5.5. Occurrence of RNA 16 in monomeric RNP particles

Some low molecular weight RNA species are believed to be components of nuclear ribonucleoprotein complexes (Raj et al., 1975; Sekeris and Niessing, 1975). In collaboration with Dr. Niessing we tried to isolate RNP particles from Friend cells labeled for 2.5 h with ^3H -uridin. In spite of the use of aurintricarboxylic acid (ATA) as an RNase inhibitor we were not able to isolate any complexes larger than 30 S monomers.

Figure 10 Separation of RNA released from poly(A)-containing RNA by gel electrophoresis according to the method of Fradin et al. (1975).

Friend cells were labeled for 18 h with (^{32}P) inorganic phosphate and the total cell RNA was centrifuged through a 5% - 23% sucrose gradient without prior heat denaturation. The gradient was divided into three pools as follows: (I) RNA $\geq$ 28 S (II) RNA between 18 S and 28 S (III) RNA < 18 S. The three preparations were applied separately to oligo(dT) cellulose in 1 x TNE and the poly(A)-containing RNA of each fraction was eluted in H_2O . The eluted material was heat denatured for 2 min at 90°C and passed through a second oligo(dT) cellulose column. The flowthrough fractions of the second column were analyzed by electrophoresis on a 1-dimensional PAA gel (1st dimension of the gel system described by Fradin et al., 1975). a) RNA released from the poly(A)-containing RNA $\geq$ 28 S b) RNA released from the poly(A)-containing RNA between 18 S and 28 S c) RNA released from the poly(A)-containing RNA < 18 S d) non-poly(A)-containing RNA < 18 S.

Fig. 10



The RNA contained in these particles was isolated and was electrophoresed on a PAA gel with a preparation of (^{32}P) RNA 16 in the neighboring lane serving as a marker (Fig. 11). Several bands with the same approximate mobility as RNA 16 can be recognized. The smear above the bands must correspond to some other RNP-bound species as reported by Raj et al. (1975) and Deimel et al. (1977). Since we did not succeed in suppressing ribonucleases, it was not possible to reach a definite conclusion and the work was not carried on further.

3.6. Relationship of RNA 16 to Z-pBR322/Mchr β G-36 DNA

The plasmid containing the mouse β -globin gene and surrounding regions described in the Materials section must have some sequences homologous to the small RNA from Friend cells since it hybridizes to it. In order to locate the region of homology with respect to the globin gene, the restriction map of the DNA was partially established and the binding site of the RNA to the DNA was determined by hybridization experiments. In order to determine whether the Z-pBR322/Mchr β G-36 was carrying a gene coding for RNA 16, the sequence of the fragment hybridizing the RNA was analyzed according to the method of Maxam and Gilbert (1977).

3.6.1. Restriction analysis of the DNA beyond the 3' end of the β -globin gene

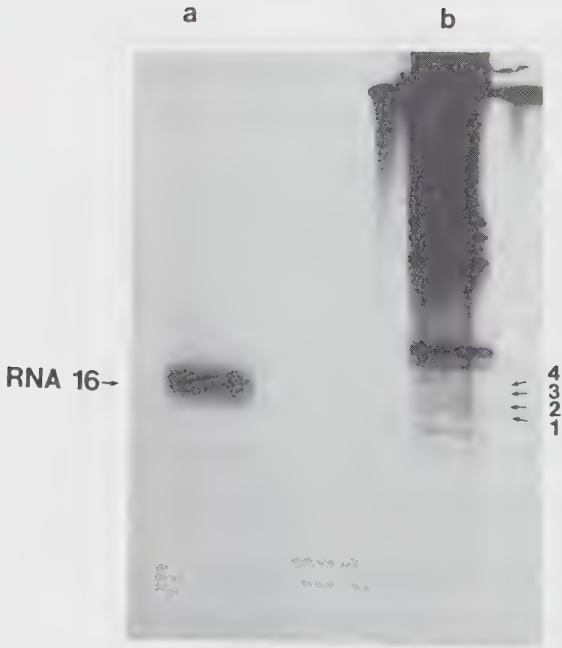
The restriction map of the BspI and the TaqI sites on the mouse β -globin gene fragment shown in Fig. 12a was established by N. Mantel, using the method of Smith and Birnstiel (1976). The restriction sites around the region of hybridization are also shown in this figure. They were determined by complete

Figure 11 PAA gel electrophoresis of small RNA from RNP particles.

Friend cells (9×10^7 cells in 25 ml medium) were labeled for 2.5 h with 5 mCi (^3H) uridine. The cells were lysed in 5 ml 50 mM Tris HCl pH 7.5, 2 mM MgCl_2 , 1 mM ATA. The lysed cells were centrifuged at 5000 rpm, the pelleted nuclei were resuspended in 1 ml extraction buffer (50 mM Tris HCl pH 7.5, 2.5 mM EDTA, 300 mM NaCl, 1 mM ATA) and sonicated six times for 3 sec at 0°C and 100 W. The nuclear suspension was centrifuged at 10'000 revs/min for 10 min. The supernatant obtained was centrifuged through a 15% - 30% sucrose gradient for 18 h at 20 K and 0°C . The material corresponding to a size of 30 S was pooled, extracted with phenol, equilibrated with 1 x TNE and separated on a denaturing PAA gel according to Fradin et al. (1975). Prior to autoradiography the gel was subjected to fluorography with PPD and DMSO.

a) Friend cell RNA 16 purified by hybridization to Z-pBR322/Mchr8G-36 DNA b) Small RNA from RNP particles.

Fig. 11



restriction endonuclease cleavages and partially confirmed by the sequence determination discussed in 3.6.5. Fig. 12b shows the pattern obtained by electrophoretic resolution of the 1100 bp BspI fragment after cleavage with restriction endonuclease AluI followed by terminal labelling with polynucleotide kinase. The size of the two AluI/BspI end fragments of approximately 35 bp and 262 bp was determined by digesting terminally labeled 1100 bp BspI fragment with restriction endonuclease AluI and characterization of the labeled fragments by gel electrophoresis. The orientation of the 1100 bp BspI fragment had to be determined in order to know the respective location of the 35 bp and 262 bp AluI/BspI fragments. This could be deduced from the fact that the 262 bp AluI/BspI fragment contains a TaqI site a few nucleotides away from the BspI end. This result, as well as the knowledge of the restriction map of the BspI and TaqI sites (Fig. 12a) indicate that the 262 bp AluI/BspI restriction fragment has to be located at the 3' end of the 1100 bp BspI fragment. The AluI sites in the 900 bp BspI fragment were determined by comparing an AluI restriction endonuclease digest of the 900 bp BspI fragment with a BspI digest of the AluI fragments. The exact position of the HinfI sites contained in the 262 AluI/BspI fragment was determined by sequence analysis.

3.6.2. Localization on Z-pBR322/Mchr8G-36 DNA of the sequence related to RNA 16 by Southern hybridization experiments

In order to define the region on Z-pBR322/Mchr8G-36 DNA responsible for the hybridization of RNA 16, a blotting experiment was carried out. In a first experiment done by N. Mantei,

Figure 12a) Restriction map of the globin insert of the Z-pBR322/Mchr8G-36 plasmid.

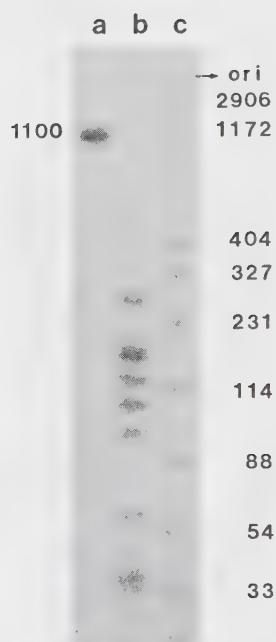
Konkel et al. (1978) determined the nucleotide sequence of the insert starting 70 nucleotides before the capping site and extending 231 nucleotides 3' of the putative poly(A) addition site. The position of the restriction sites extending to either side of this region was only determined by restriction mapping and is therefore not precise. Only the AluI sites and the HinfI sites important for the location and sequence determination of the DNA stretch hybridizing RNA 16 were analyzed.

The size indicated for the end - standing BspI and TaqI fragments includes PBR sequences up to the next BspI respectively TaqI site.

Figure 12b) Determination of the size of the AluI fragments contained in the 1100 bp BspI fragment.

The 1100 bp BspI fragment of the mouse chromosomal β -globin plasmid DNA was digested with restriction endonuclease AluI, the fragments were terminally labeled with polynucleotide kinase in presence of (32 P) γ -ATP and separated by electrophoresis on a 5% PAA gel with TaqI digested ϕ X 174 DNA in the neighboring lane. a) 1100 bp BspI fragment, b) 1100 bp BspI fragment x AluI, c) ϕ X 174 DNA x TaqI.

Fig. 12b



the DNA was cleaved by the restriction enzymes BspI, TaqI and AluI, separated on agarose gel (Fig. 13a) and transferred to nitrocellulose filters according to Southern (1975). In vivo labeled, filter-purified (^{32}P) RNA 16 was shown to hybridize to the 1100 bp BspI fragment, the 1200 bp TaqI fragment as well as to an AluI fragment of a size between 150 -300 bp. According to the restriction map these results indicate a location of the hybridizing site in an area about 1400 bp - 2500 bp downstream from the 3' end of the globin gene. In order to determine more accurately the size of the AluI fragment hybridizing RNA 16, a second Southern experiment was carried out. Z-pBR322/Mchr8G-36 DNA was digested with restriction endonuclease AluI and separated by electrophoresis. pBR322 cleaved with the enzymes AluI or BspI were run in parallel lanes. The size markers allowed to locate the size of the hybridizing AluI fragment between 250 and 280 bp. As shown in Fig. 12b, the 1100 bp BspI fragment is cleaved by AluI into 9 subfragments ranging between 262 and 35 bp, indicating that the largest of these 9 fragments is the only possible candidate. The result of this experiment is shown in Fig. 13b.

3.6.3. Frequency of sequences related to the AluI/BspI 262 bp fragment in the total mouse genome

Analysis of cloned genes in several eucaryotic species had indicated the presence of middle repetitive DNA sequences dispersed throughout the "unique" sequences in the genome. In order to obtain an estimate of the number of copies of DNA homologous to the 262 bp AluI/BspI fragment, samples of total mouse DNA were digested in parallel with restriction endonu-

Figure 13 Localization on Z-pBR322/Mchr6G-36 DNA of the sequence related to RNA 16 by Southern hybridization experiments.

a) Plasmid DNA was digested with the restriction endonucleases TaqI, AluI or BspI and the fragments were separated on a 1.8% agarose gel in Tris-acetate buffer. The gel was stained with EtBr, a picture was taken and the gel was treated with alkali according to Southern (1975). After the transfer, the hybridization was carried out with in vivo labeled (^{32}P) RNA (prepurified by filter hybridisation followed by gel electrophoresis as described in the Legend to Figure 2) in 50% formamide and 0.5 M NaCl for 60h at 37°C . 1) photograph of the EtBr stained gel showing the pattern of the TaqI digest of the mouse plasmid, 2) autoradiograph of the TaqI digested DNA after hybridization to ^{32}P labeled RNA from band 5S-6 (corresponding to RNA 16, Figure 2) 3) photograph of AluI digested DNA, 4) autoradiograph of (3) after the hybridization, 5) photograph of BspI digested DNA, 6) autoradiograph of (5) after the hybridization. 2'), 4') and 6') are as 2), 4) and 6) except that the (^{32}P) labeled RNA used for hybridization corresponds to species 5S-4 (Figure 2).

b) Method as in Fig. 13a. 1) photograph of pBR322 DNA digested with restriction endonuclease BspI,

Figure 13 (cont.)

2) photograph of pBR322 DNA digested with restriction endonuclease AluI, 3) photograph of Z-pBR322/Mchr8G-36 DNA digested with restriction endonuclease AluI, 4) autoradiograph of 3) digested with Alu I after hybridisation to (^{32}P) labeled RNA 16.

Fig. 13a

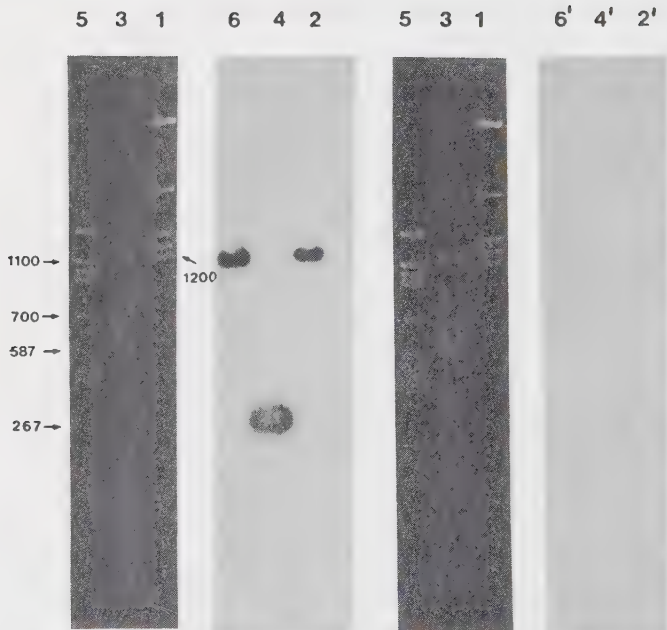
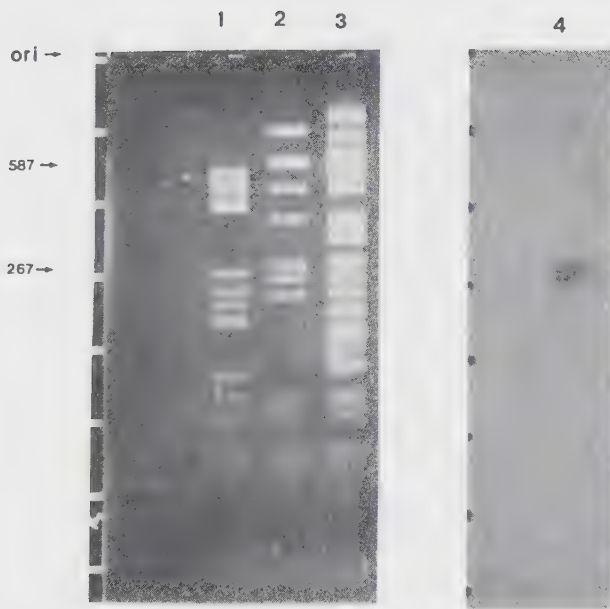


Fig. 13b



cleases EcoRI or PstI and used in a Southern type blotting experiment to hybridize nick-translated 262 bp fragment. For calibration Z-pBR322/Mchr β G-36 DNA (1,2 and 5 copies of genome equivalent) was cleaved with the restriction endonucleases EcoRI and PstI and applied in parallel to the same gel. As shown in Fig. 14, the sequence related to the 262 bp AluI/BspI fragment has a frequency of several thousand copies per haploid genome.

3.6.4. Identification of the strand hybridizing to RNA 16

The result obtained by Southern hybridization (Section 3.6.2.) was confirmed by liquid hybridization experiments using single-stranded DNA fragments. For this purpose, the terminally labeled 262 bp AluI/BspI fragment was heat denatured and its complementary strands were separated on a 7% PAA gel. The single-stranded fragments were each hybridized in liquid to an excess of the 4 - 5S fraction of Friend cell RNA. As shown in Fig. 15, the slower running single strand hybridizes to the RNA, the slower migrating material corresponding to hybrid molecules of larger molecular weight. The appearance of various bands of different mobility might be explained by the presence of some other, larger RNA species (see Section 3.3.) which probably contain sequences partially related to RNA 16. Another explanation could be that these different bands correspond to various types of secondary structures in the RNA and DNA. The possible secondary structures of the two molecules will be discussed in the last paragraph. Since we wanted to know the exact location of the hybridized RNA on the DNA fragment, the DNA was further subdivided by restriction cleavage. Digestion of the

Figure 14 Autoradiograph of a Southern blotting experiment of total mouse genomic DNA using nick translated 262 bp AluI/BspI fragment as a probe.

Mouse genomic DNA (10 μ g) was digested with endonuclease EcoRI or PstI respectively and run on a 0.85% agarose gel in Tris acetate pH 7.8 containing 0.5 μ g/ml EtBr. Z-pBR322/Mchr β G-36 DNA, 1, 2 and 5 copies of genome equivalent (47 pg, 94 pg and 235 pg), was also digested with the endonucleases EcoRI or PstI and run on the same gel. The gel was treated twice for 15 min with 0.25 M HCl, rinsed extensively with water, soaked twice for 15 min in 0.5 M NaOH, 1M NaCl, rinsed again with water and finally neutralized by repeated washing with 0.5 M Tris pH 7.5, 3 M NaCl. The DNA was transferred to a Sartorius filter soaked in 20 x SSC at room temperature during 3.5 h. 100 ng of the 262 bp AluI/BspI DNA fragment was labeled (1×10^8 dpm) by nick translation according to Rigby et al. (1977) and hybridized in 50% formamide, 5x Denhardt mixture (Denhardt, 1966), 5 x SSC and 100 μ g/ml alkali treated salmon sperm DNA for two days at 37 $^{\circ}$ C. The filter had previously been prehybridized for 4 h in the same buffer. a) 47 pg plasmid DNA x EcoRI, b) 94 pg plasmid DNA x EcoRI, c) 235 pg plasmid DNA x EcoRI, d) 10 μ g total mouse genomic DNA x EcoRI, e) 10 μ g total mouse genomic DNA x PstI, f) 47 pg plasmid DNA x PstI, g) 94 pg plasmid DNA x PstI, h) 235 pg plasmid DNA x PstI

Fig. 14

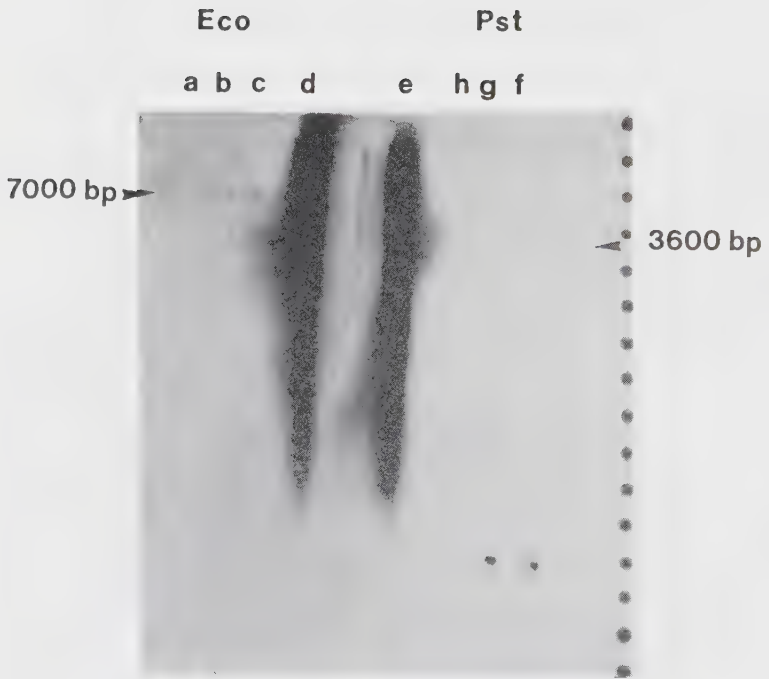


Figure 15 Hybridization of low molecular weight crude Friend cell RNA to the terminally labeled separated strands of the 262 bp AluI/BspI fragment.

5' terminally labeled 262 bp AluI/BspI fragment was heat denatured and the strands were separated on a 7% PAA gel in 89 mM Tris borate buffer. The single-stranded DNA fragments were eluted out of the gel with 5 ml 150 mM NaCl, 50 mM Tris HCl pH 8.0, 2 mM EDTA and concentrated on a DEAE column. The DNA was eluted from the column with 0.4M NaOH, the eluant was neutralized with 100% acetic acid and precipitated in ethanol. The separated strands were then individually hybridized in solution to cold excess RNA from the 4 - 5 S pool of a sucrose gradient centrifugation in 50% formamide, 0.75 M NaCl, 50 mM Hepes pH 7.0, 0.5% SDS and 5 mM EDTA. The hybridization was carried out for 6 h at 45°C, followed by 18 h at 37°C. The samples were electrophoresed on a 6.6% PAA gel (as in Legend to Fig. 2). a) fast migrating strand, b) self annealed fast migrating strand, c) fast migrating strand after hybridization to the RNA, d) double-stranded 262 bp DNA fragment, e) slow migrating strand after hybridization to the RNA, f) self annealed slow migrating strand, g) slow migrating strand.

Figure 16 Hybridization of low molecular weight crude Friend cell RNA to the terminally labeled separated strands of the 26 bp and 82(79) bp HinfI fragments.

Terminally labeled HinfI DNA fragments were individually heat denatured and the strands were separated by PAA gel electrophoresis. The single-stranded DNA fragments were eluted out of the gel with 5 ml 150 mM NaCl, 50 mM Tris HCl pH 8.0, 2 mM EDTA and concentrated on a DEAE column. The DNA was eluted from the column with 0.4 M NaOH, the eluant was neutralized with 100% acetic acid and precipitated in ethanol. Each preparation was hybridized as indicated in the legend to Fig. 15 to cold excess crude RNA from the 4 - 5S gradient fractions. The resulting samples were run on an 8% PAA gel.

a) 26 bp fragment, fast strand hybridized to RNA, b) 26 bp fragment, fast strand, c) 26 bp fragment, slow strand hybridized to RNA, d) 26 bp fragment, slow strand, e) 82(79) bp fragment, slow strand hybridized to RNA, f) 82(79) bp fragment, slow strand, g) 82(79) bp fragment, fast strand hybridized to RNA, h) 82(79) bp fragment, fast strand.

Fig. 15

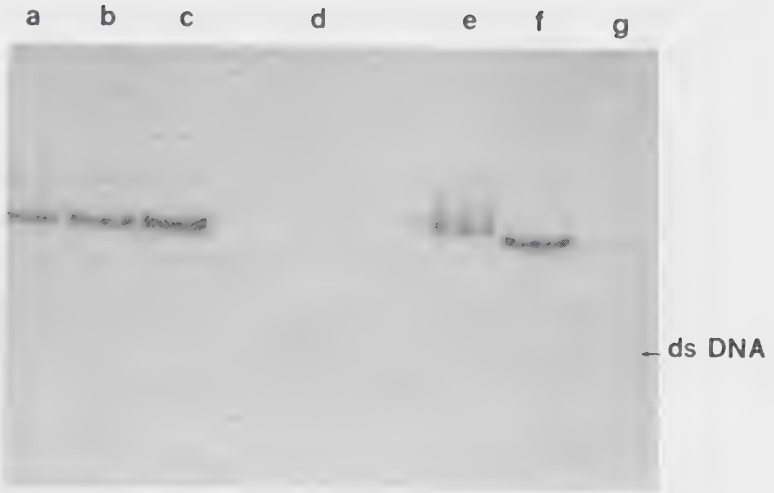
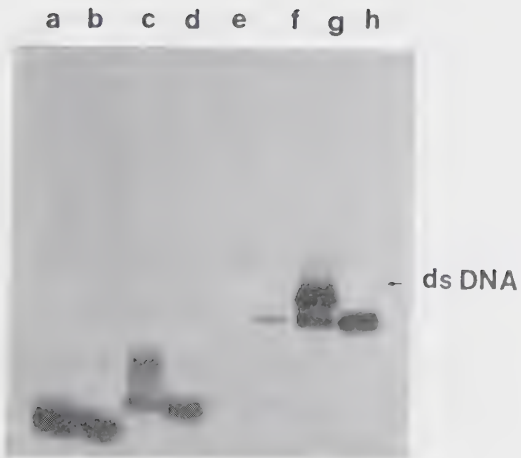


Fig. 16



fragment with restriction endonuclease HinfI yielded four fragments of 48(51), 26, 106 and 82(79) base pairs (lengths based on sequencing). The separated strands of each fragment were hybridized to unlabeled Friend cell RNA as described for the 262 bp AluI/BspI fragment. The result of this experiment is shown in Fig. 16. Friend cell RNA changes the mobility of two sets of fragments. One strand of the 26 bp fragment yields three bands, two of them of lower mobility than the starting material. The presence of two bands of lower mobility can be explained by different secondary structures. Furthermore, since the RNA used for the hybridization is a mixture of all species of 4 - 5S in the cell, the presence of another RNA species carrying a small region homologous to the 26 bp fragment is not excluded. In the case of the 82(79) nucleotide fragment, both strands show a lower mobility. The change of mobility of the fast strand is accounted for by an extensive sequence complementarity between the DNA and the RNA. As indicated in Fig. 16, one band is due to reannealing to double-stranded DNA because of cross-contamination from the other strand. The sequencing data of the DNA do not support the hybrid formation of RNA 16 with the slow strand of the 82(79) bp fragment. A possible interpretation is the formation of triple hybrids because of the cross-contamination from the other strand. It is also possible that another RNA species present in the crude Friend cell 4 - 5S fractions carries a sequence homology with this strand.

3.6.5. Determination of the nucleotide sequence of the region of plasmid Z-pBR322/McchrBG-36 involved in binding RNA 16

The sequence of the 262 bp AluI/BspI fragment shown in Fig.

Figure 17 Nucleotide sequence of a segment of Z-pBR322/Mchr-
βG-36 DNA extending 262 bp 5' and 31 bp 3' of the
BspI site (pos. 2500 on Map on Fig. 12a)
Arrows indicate the extent of the sequence deter-
mination from each experiment. Possible inaccura-
cies in the sequence determination are marked with
stars (☆) and nucleotides which could not be deter-
mined are indicated with an X.

■ - - -> Sequencing experiment using strand-separated
DNA. The tip of the arrow indicates the last
nucleotide which could be read.

▶▶ Sequencing experiment using double-stranded
DNA labeled at one end. The tip of the arrow
indicates the last nucleotide that could be
read.

Fig. 17

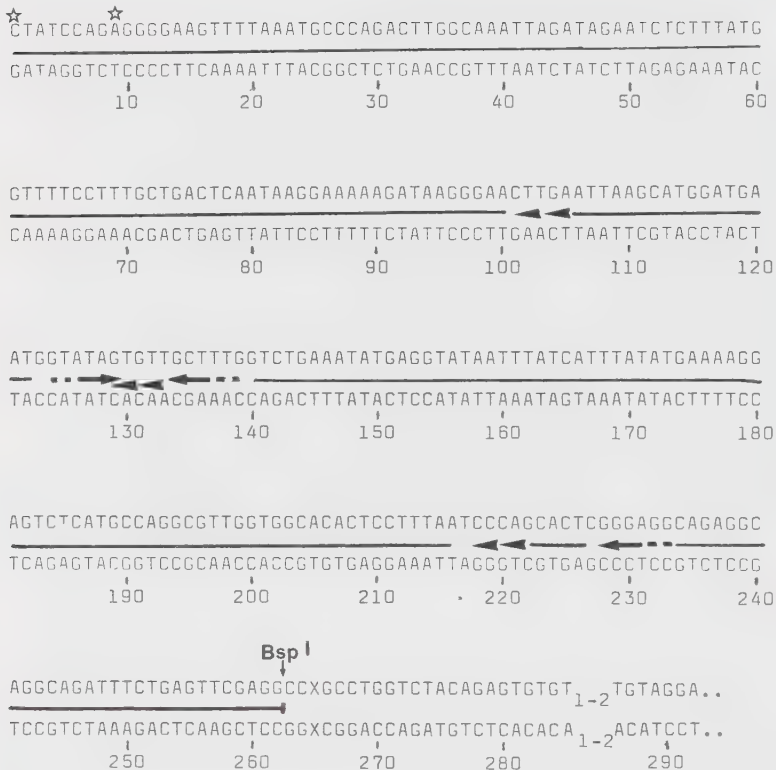
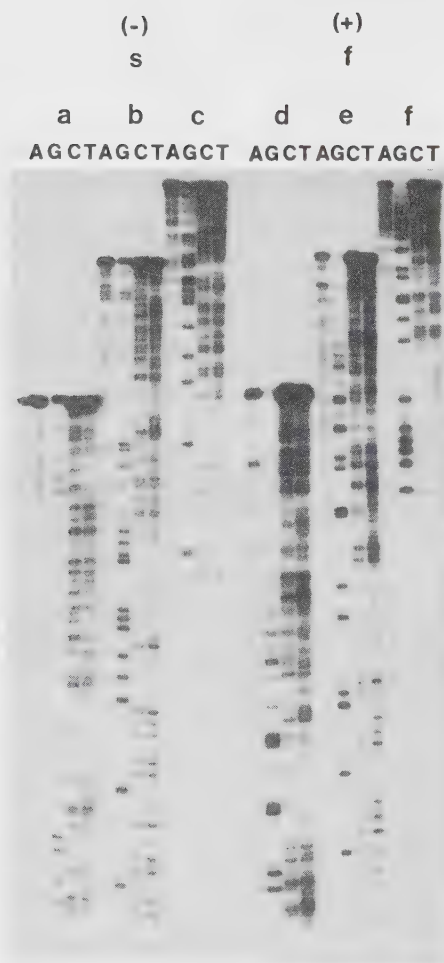


Figure 18 Representative autoradiograph of a sequencing gel.

262 bp AluI/BspI fragment was 5' terminally labeled with polynucleotide kinase and the two strands (slow, fast) were separated by gel electrophoresis as described in the legend to Fig. 15. Each strand was subjected to sequence analysis using the method of Maxam and Gilbert (1977). All runs were carried out on a 12% gel. a) 17 h run of the s strand, b) 7 h run of the s strand, c) 2.5 h run of the s strand, d) 17 h run of the f strand, 3) 7 h run of the f strand, f) 2.5 h run of the f strand.

Fig. 18



17 was determined as follows: In a first approach, the 1100 bp BspI fragment was digested with AluI, the resulting fragments were 5' terminally labeled and the band corresponding to the 262 bp AluI/BspI fragment was isolated by PAA gel electrophoresis. Its complementary strands were separated as described in the legend to Fig. 15 and the sequence of each single-stranded DNA fragment was determined according to the method of Maxam and Gilbert (1977). In a second experiment, the 1100 bp BspI fragment was 5' terminally labeled and subsequently digested with AluI and the sequence of the resulting 262 bp AluI/BspI fragment (labeled at the BspI end) was determined. A representative autoradiograph of a sequencing gel is shown in Fig. 18. The sequence of the 5' end of the 900 bp BspI fragment was only determined once and contains some ambiguities. A comparison of the sequenced DNA fragments with RNA 16 shows immediately that RNA 16 is strongly homologous to the (+) strand of the DNA, however that there are enough differences to exclude this DNA segment as a template for the synthesis of RNA 16. The DNA itself has several interesting features. There is a rather frequent occurrence of termination codons in both strands and in all three frames. In addition the fragment contains a sequence homologous to the so-called Alu family, a reiterated DNA sequence found in various mammalian genomes (Jelinek et al., 1980). It is also strongly homologous to a cloned DNA B₁ sequence, corresponding to the most abundant class of mouse foldback RNA (Krayev et al., 1980). The relationship between these groups of sequences is shown in Fig. 19.

Figure 19a Comparison of the nucleotide sequence of DNA and RNA fragments related to the Alu family of reiterated sequences.

Figure 19b Homology between the nucleotide sequence of RNA 16 and the 262 bp AluI/BspI fragment with the B₁ type sequence by Krayev et al. (1980) and other related sequences.

*indicate the nucleotides identical to the consensus B₁ sequence by Krayev (1980).

Fig. 19 a

a	CXACTCGGGAGGCTGAGGCAGAGAAATCGCATGAACC	Alu family consensus
b	CTACTTAGGAGGCTGAGACAGAGAATCCCTTAACCC	BLUR 8
c	CTACTTGGAGGCTGAGGAGGAGGAATTGCTTAAACC	Pol III ^G -globin template
d	GCACTCGGGAGGCAGAGGCAGGCAGATTCTCTGAGTT	Z-pBR322/Mchr8G-36 (+) strand
e	UGAAGGAGGCACAGGCAGGAGGAUCACGA	CHO sn RNA
f	UGAAGGAGGCACAGGCAGGGAUCACGA	RNA 16
g	GGGTGGAGGCACAGGCGGCTC	BK Virus Ori
h	GCTGAGAGGCGGAGGCGGCTC	SV40 Ori

Footnote:

All except sequence (d) and (f) are taken from Jelinek et al. (1980).
(f) was determined by Harada and Kato (1980).

Fig. 19 b

```

1  CCGGGCA TGG TGGTGCATGCCTTTATTCCCAGC AC TCGGGAGGCAGAGGCAGGCGGATTTCTG
2  GCCAGGCGTTGG TGGCACACTCCTTTAATCCCAGC AC TCGGGAGGCAGAGGCAGGCAGATTTCTG
3  GCC GGUAGUGG UGGCGCACGCCGGUAGGAUUG CU GAAGGAGGCAGAGGCAGAGGGAUAC G
4  GGC TAGGCGCGGGTTCACGCCTGTAATCCCAGC ATTTTGGGAGGCTGAGACGGGTGGATCATGA
5  .CCGAGTGTGG TGGTGGGCACCTGTAGTCCCAGCTAC TCAGGAGGCTGAGGCAGGAGAATGACTT
6  GCATG GA GACGGAGGCAGAGGTGGGCAGATCATGA
7  CTGGGCA TGG TGGCATGCCTGTAGTCCCAGCTGC TCGGGAGGCTGAGGCAGGAGAATCGTTT

```

```

1  AGTTCGAGGCCAGCCTGGTCTTCAGAGTGTAGTTCC AGGACACCAGGGCTACAGAGAAACCCTGTCT
2  AGTTCGAGGCCXGCCCTGGTCTACAGAGTGT GTTAGGA
3  AGUUCGAGGCCAGCCUGGGCUACACA2-5UOH
4  GGTACGAGATCGAGACCATCT...
5  GAACCTGGAGGTGGAG CTTG CAGTGAGCCAAC
6  GGTCAAGAGATCG
7  GAACCCAGGAGGCGAAGGTTG CAGTGAGCTGAGATAGTGCCATTGCACTCCAGTCTGGCAACAG

```

- 1 B₁ consensus sequence (Krayev et al., 1980)
- 2 Z-pBR322/Mchr3G-36 (+) strand DNA
- 3 RNA 16 (Harada et and Kato, 1980)
- 4 and 5 polymerase III template in human DNA (Alu family members)
from Elder et al. (1981)
- 6 and 7 5' flanking region of the human ε-globin gene (Alu family
members) from Baralle et al. (1980)

3.7. Structure of the DNA-RNA hybrid

In order to demonstrate experimentally the structure of the partial hybrid formed between RNA 16 and the segment of Z-pBR322/Mchr8G-36 DNA, partially homologous to it, the following three approaches were used.

3.7.1. RNase T₁ treatment of the DNA-RNA hybrid

In vivo (^{32}P)-labeled RNA 16 was hybridized to filter-bound Z-pBR322/Mchr8G-36 DNA and the structure of the partial hybrid was analyzed by treating it with ribonuclease and determining the oligonucleotide composition of the resistant and sensitive material by fingerprint analysis. As shown in Fig. 20a the resistant material, when separated by PAA gel electrophoresis, gave one main band and two fainter bands. The bands were fingerprinted in parallel with the RNase T₁ sensitive material. For most of the RNA (band no. 2 on the PAA gel), the protection extends from position 37 to the 3' end of the molecule (20b+c). The band of lowest mobility (no. 1) represents material containing eight additional protected nucleotides, whereas the RNA fragment in band no. 3 corresponds to a protection starting at position 47. In conclusion this approach shows that the strongest interaction between RNA 16 and the plasmid DNA occurs at the 3' half of the molecule. The regions of complementarity near the 5' end of the RNA seems not to be extensive enough to afford stable hybrid structures under these digestion conditions. It is not clear however, why the RNase does not cleave the G residues inside the loop corresponding to positions 54 - 56 on the RNA, yielding an even shorter resistant fragment extending from position 57 to the 3' end of the RNA. The presence of spot no. 5 in the fingerprint of

Figure 20 Analysis of the fraction of (^{32}P)-labeled RNA 16 released from a Millipore filter containing Z-pBR322/Mchr β G-36 DNA by digestion with RNase T_1 .

(^{32}P)-labeled RNA 16, purified by preparative hybridization to a Millipore filter containing Z-pBR322/Mchr β G-36 DNA, was rehybridized to the same DNA bound to a filter as described in the Methods section. After the hybridization the filter was rinsed with 0.5 M NaCl, 10 mM Tris HCl pH 7.5. RNase T_1 treatment of the filter was carried out in 1 ml of the same solution, containing 20 U of RNase T_1 and 50 μg yeast carrier RNA for 30 min at 37°C . The material released by RNase T_1 was subjected to fingerprint analysis. The filter was rinsed twice with 0.5 M NaCl, 10 mM Tris HCl pH 7.5, the buffer was adjusted to 1% SDS + the filter was treated for 15 min at 37°C with 10 μl pronase (20 mg/ml) in 1 ml of buffer. The filter was extracted twice with phenol, the RNA was eluted from the filter with formamide as described in the Methods section and fractionated by PAA gel electrophoresis. The three RNA fragments were extracted, digested to completion with RNase T_1 and the oligonucleotides separated by fingerprinting.

20 a) PAA gel electrophoresis of the RNA fragments resistant to RNase T_1

Figure 20 (cont.)

20 b) T_1 fingerprint analysis of the RNA oligonucleotides resistant and sensitive to the digestion with RNase T_1 :

- 1) T_1 fingerprint of the resistant material from band 1 in Fig. 20 a.
- 2) T_1 fingerprint of the resistant material from band 2 in Fig. 20 a.
- 3) T_1 fingerprint of the resistant material from band 3 in Fig. 20 a.
- 4) T_1 fingerprint of the material released by RNase T_1 digestion.

20 c) Sequence scheme showing the interpretation of the results from the fingerprint analysis on one of the most probable hybrid models (III).
(All the models for the DNA - RNA hybrids and the schemes for the secondary structures of the single - stranded molecules are shown in the Appendix.)

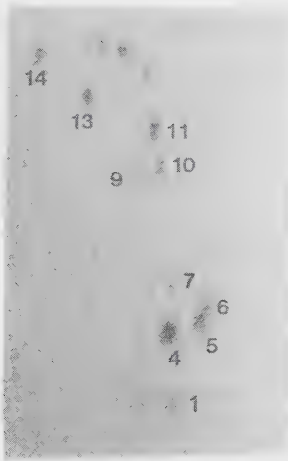
Fig. 20a

- 1-
- 2-
- 3-



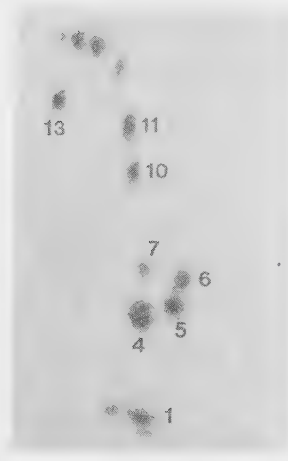
Fig. 20b

I



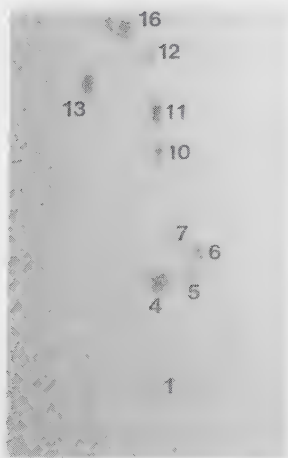
band 1 product

II



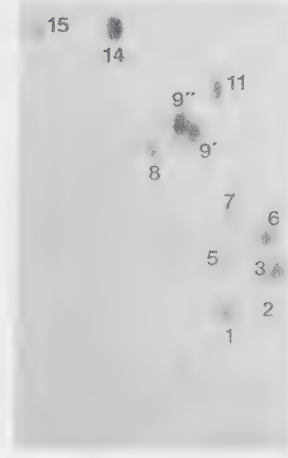
band 2 product

III



band 3 product

IV



eluted material

the material from band no. 3 (the band of highest mobility) however speaks for the fact that this band corresponds to material starting at position 47 and extending to the 3' end of the RNA.

3.7.2. Partial degradation by nuclease S_1 of the hybrid using terminally (^{32}P)-labeled DNA

For this experiment a hybrid was made with RNA 16 using 5' terminally labeled 262 bp AluI/BspI DNA fragment. In an experiment inspired by the footprinting method by Galas and Schmitz (1978), we tried to differentiate between the areas of the DNA protected from nuclease S_1 digestion by the hybridization to the RNA and the regions exposed to nuclease attack. The DNA-RNA hybrid was treated under very mild conditions with nuclease S_1 , in order to get an average of 1 cleavage per DNA strand. The samples were analyzed by electrophoresis on a sequencing gel (Maxam and Gilbert, 1977), allowing the visualization of the distribution of the cleavage sites. Preferred cleavage sites (i.e. nonhybridized regions) should give rise to strong bands of DNA fragments of the appropriate size class, whereas bands corresponding to DNA segments protected by the RNA should not be seen. As a control experiment, single-stranded DNA was digested under the same conditions as the hybrid. The nuclease S_1 cleavage patterns are shown in Fig. 21a and 21b. Fig. 22 shows an interpretation of the deduced cleavage sites on the hypothetical hybrid structures and on the secondary structure of the (-) strand DNA fragment. The cleavage pattern in case of the single-stranded DNA fragment only partially supports the model for the most probable secondary structure obtained by computer analysis of the DNA sequence. Except for a low background covering most of the sites, all cleavage sites

Figure 21 Partial S_1 cleavage experiment of RNA 16 hybridized to the (-) strand 262 base AluI/BspI fragment.

RNA 16 was hybridized in excess to terminally labeled 262 base AluI/BspI single strand as described in the legend to Fig. 15. The hybrid was subjected to treatment with increasing amounts of S_1 nuclease in 100 μ l 0.25 M NaCl, 30 mM NaOAc pH 4.5, 1 mM $ZnSO_4$ at 25°C for 40 min. As a control experiment, single-stranded DNA was digested under the same conditions. The digestion products were analyzed on a 12% sequencing gel according to Maxam and Gilbert (1977). In Fig. 21a the gel was run at 900 V for 2.5 h, in Fig. 21b for 7 h.

a) lanes 1 and 19, pBR322 plasmid DNA digested with restriction endonuclease BspI; lane 2, DNA-RNA hybrid undigested; lanes 3-9, DNA-RNA hybrid digested with 0.01 U, 0.02 U, 0.04 U, 0.08 U, 0.12 U, 0.20 U, 0.30 U of S_1 nuclease, respectively; lane 10, single-stranded DNA undigested; lanes 11-17, single-stranded DNA digested with 0.01 U, 0.02 U, 0.04 U, 0.08 U, 0.12 U, 0.20 U, 0.30 U of S_1 nuclease, respectively; lane 18, C + T degradation pattern of single-stranded DNA according to the sequencing method by Maxam and Gilbert (1977).

b) lanes 1 and 19, G sequencing pattern of single-stranded DNA; lane 2, C + T sequencing pattern of

Figure 21 (cont.)

single-stranded DNA; lane 3, DNA-RNA hybrid undigested; lanes 4-10, DNA-RNA hybrid digested with 0.01 U, 0.02 U, 0.04 U, 0.08 U, 0.12 U, 0.20 U, 0.30 U of S_1 nuclease, respectively; lane 11, single-stranded DNA undigested; lanes 12-18, single-stranded DNA digested with 0.01 U, 0.02 U, 0.04 U, 0.08 U, 0.12 U, 0.20 U, 0.30 U of S_1 nuclease, respectively; lane 20, pBR322 plasmid DNA digested with restriction endonuclease BspI.

Fig. 21a

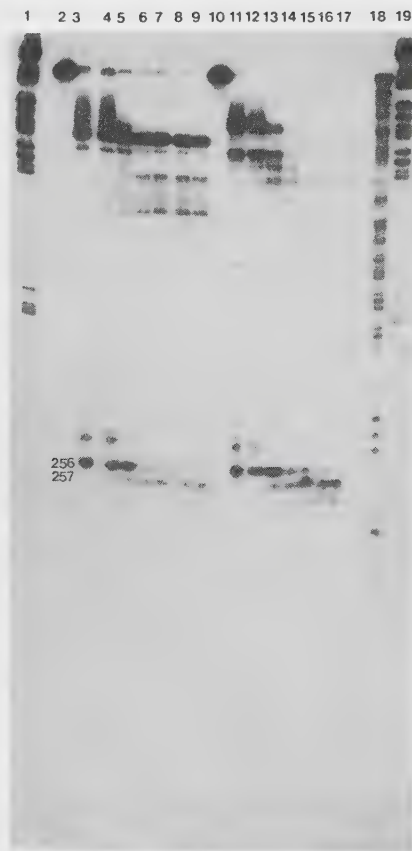


Fig. 21b

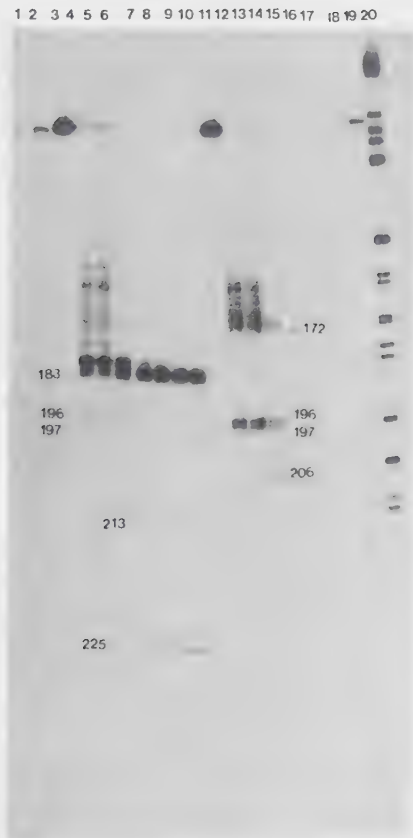


Figure 22 Most probable models for the structure of the DNA-RNA hybrid between RNA 16 and the (-) strand of the 262 base long AluI/BspI fragment and hypothetical secondary structure of the portion of the DNA normally involved in binding the RNA.

Arrows indicate the sites of cleavage of the S_1 nuclease in the experiment described in Fig. 21 a and b.

are located in hypothetical single-stranded regions. It is interesting to note however that S_1 nuclease does not cleave the single-stranded areas in a strictly random fashion, even though it is known to have a very limited base specificity regarding the site of nuclease activity (Vogt, 1973). This result could be interpreted in several ways. Even though no other striking complementary regions involving the undigested nucleotides are found by computer analysis, the secondary structure of the molecule could be different than expected. Several groups have used S_1 nuclease to identify the conformation of RNA molecules such as tRNAs (Wrede et al., 1979) or small nuclear RNAs (Branlant et al., 1981). It was observed that not all single-stranded positions of the RNA were equally well digested with nuclease S_1 . In fact, some regions known to be single-stranded were not cleaved at all. Since nuclease S_1 is a relatively large enzyme (35'000 as compared to 11'000 for T_1 nuclease) this finding may be explained by steric hindrance due to the tertiary structure of the DNA molecule. In the case of the DNA-RNA hybrid, the sites of preferential cleavage are located at the end of the putative double-stranded region, around position 183 to 188 on the DNA (-) strand. About 20 - 30 nucleotides 3' of this region are also cleaved to a lower extent. The other cleavages (pos. 196, 197, 213 and 225) are located on hypothetical loop structures. The strong cleavages located close to the labeled end, in the same position as in the single-stranded molecule are not in agreement with models I, II, III or IV but rather with model V, where the 3' region of the RNA contains a secondary structure rather than forming a hybrid with the DNA (models in Appendix).

3.7.3. Footprinting experiment using sodium bisulfite as a single-strand specific reagent

Sodium bisulfite has been shown by many workers to produce uracil residues by deamination of cytosine residues (Shapiro et al., 1973). With the use of an infectivity test, Shortle and Nathans (1978) have underlined the single-strand specificity of the reagent which should make it a tool for differentiating between single-stranded and hydrogen-bonded regions on a nucleic acid molecule. The partial hybridization of an RNA molecule to a single-stranded DNA fragment should protect the regions of the DNA which are base-paired to the RNA from the effect of bisulfite, whereas the bisulfite should convert all cytosine residues into uracil residues on the non-protected single-stranded regions of the DNA molecule. By subsequent sequence determination of the terminally labeled DNA molecule it should be possible to differentiate between the mutagenized regions and the ones which remained untouched. The hybrid consisting of terminally labeled (-) strand of the 262 bp AluI/BspI fragment and RNA 16 was treated with sodium bisulfite using the conditions described by Shortle and Nathans (1978). A series of pilot experiments showed that the bisulfite treatment had an unexpected effect on single-stranded DNA. The subsequent hydrazine reaction in presence of salt (normal reaction conditions specific for C cleavage) gave rise to a normal C pattern, since the uracil residues seemed to react at the same rate as cytosine. However, in absence of salt (reaction conditions specific for C + T cleavage) the uracil groups reacted very fast and only very small fragments were obtained under the normal con-

ditions for the hydrazine reaction. By reducing the reaction times with the hydrazine drastically, we attempted to limit the extent of cleavage to such a degree that uracil residues are attacked almost exclusively (in the C + T cleavage). Even at the shortest reaction times however, the cleavage at every uracil residue still was extensive, resulting in a low average fragment size. For a DNA partially protected from bisulfite by basepairing to RNA the hydrazine cleavage would therefore be expected to result in a fragment pattern in which the bands in the non-basepaired segment appear strongest. The result as shown in Fig. 23a and schematically drawn in Fig. 23b indicates however, that the strongest bands correspond to cleavage sites which according to the hybridization models should be in a base-paired region. Supposedly single-stranded segments give rise to only weak bands, except for the 5' terminal region. This outcome can be partially explained by assuming that RNA binding protects the C-residues in the short basepaired regions only partially from bisulfite attack, such that the low salt hydrazine cleavage appears accelerated not only in the single-stranded, but to some extent also in the basepaired portions of the DNA. However, it is difficult to understand why the second base-paired region from the labeled end still shows such a high intensity of label after cleavage at the C-residue since most of the label should have been consumed by the preferential cleavage in the first loop. Nevertheless, the method allows to resolve features as small as the loop at position 194-197 on the DNA. Other unpaired regions, like the loop at position 248-252 cannot be resolved due to their lack of C-residues.

Figure 23a Treatment with sodium bisulfite of the DNA-RNA hybrid composed of RNA 16 and the (-) strand of the 262 base AluI/BspI DNA fragment.

Cold excess RNA from the 4-5S pool of a sucrose gradient centrifugation of Friend cell RNA was hybridized to 5' terminally labeled single-stranded DNA under the conditions described previously. The hybrid was incubated for 24 h at 37°C in 3 M bisulfite, 0.5 mM hydroquinone in the dark. To terminate the reaction, the material was first desalted on a Sephadex column and incubated overnight in 0.2 M Tris-buffer pH 9.2 at 37°C. The alkaline pH allows the elimination of the sulfite group from the deaminated intermediate. Non-hybridized DNA was treated with bisulfite under the same conditions. Lanes 1-4 correspond to the 4 sequencing tracks of the single-stranded DNA (-) strand according to the method of Maxam and Gilbert (1977); lane 5, cleavage reaction specific for T of the DNA-RNA hybrid after treatment with sodium bisulfite, hydrazine reaction time of 40 sec; lane 6, cleavage reaction specific for T of the single-stranded DNA after sodium bisulfite treatment, hydrazine reaction time of 40 sec; lane 7, cleavage reaction specific for T of the single-stranded DNA, no bisulfite treatment, hydrazine reaction time of 40 sec; lane 8, cleavage reaction specific for T of the DNA-RNA hybrid after treatment with sodium bisulfite, hydra-

zine reaction time of 90 sec; lane 9, cleavage reaction specific for T of the single-stranded DNA after sodium bisulfite treatment, hydrazine reaction time of 90 sec; lane 10, cleavage reaction specific for T of single-stranded DNA, no sodium bisulfite treatment, hydrazine reaction of 90 sec; lane 11, cleavage reaction specific for T of the DNA-RNA hybrid after reaction with sodium bisulfite, hydrazine reaction time 3 min; lane 12, cleavage reaction specific for T of single-stranded DNA after reaction with sodium bisulfite, hydrazine reaction time 3 min; lane 13, cleavage reaction specific for T of single-stranded DNA, no sodium bisulfite treatment, hydrazine reaction time 3 min; lane 14, cleavage reaction specific for T of the DNA-RNA hybrid after treatment with sodium bisulfite, hydrazine reaction time 7.5 min; lane 15, cleavage reaction specific for T of single-stranded DNA, no sodium bisulfite treatment, hydrazine reaction time 7.5 min; lane 16, cleavage reaction specific for T of single-stranded DNA, no sodium bisulfite treatment, hydrazine reaction time 7.5 min; lane 17, cleavage reaction specific for T of the DNA-RNA hybrid after treatment with sodium bisulfite, hydrazine reaction time 15 min; lane 18, cleavage reaction specific for T of single-stranded DNA after treatment with sodium bisulfite, hydrazine reaction time 15 min; lane 19, cleavage reaction specific for T of single-stranded DNA, no sodium bisulfite treatment, hydrazine reaction time 15 min.

Fig. 23a

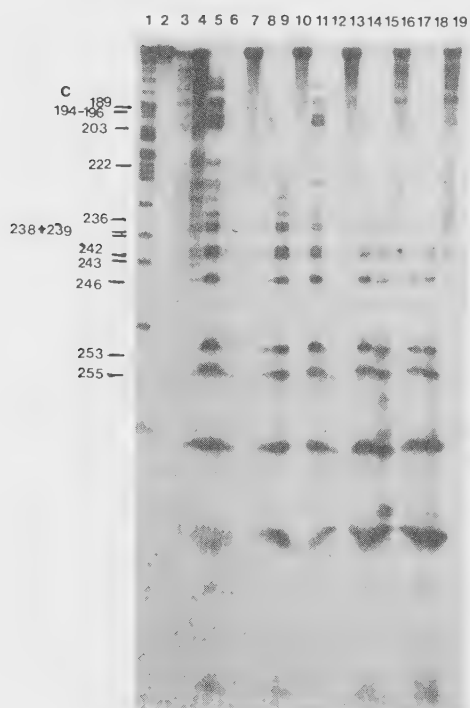


Figure 23 b Schematic interpretation of the results from the sodium bisulfite treatment of the DNA-RNA hybrid followed by the cleavage reaction after T according to the method of Maxam and Gilbert (1977).

An attempt is made to correlate the intensities of the bands appearing on the sequencing gel with one of the most probable hypothetical hybrid structures (III) between RNA 16 and the (-) strand of the 262 bp AluI/BspI fragment.

The C residues are marked with a star (★). Each arrow of the graph corresponds to one C residue.

4. DISCUSSION

4.1. The nucleotide sequence of RNA 16 and its comparison to other known small RNA molecules

The presence of a triphosphate moiety at the 5' end suggests that RNA 16 may be a primary transcript. Since the gene or the genes coding for RNA 16 have not been cloned yet it is not certain whether the oligo(U) stretch at the 3' end of the molecule is copied from the DNA sequence or whether it is added to the RNA after the end of the transcription by a poly(U) polymerase. Harada and Ikawa (1979) report that RNA 16 associated with SFV virions is composed of up to 30 different species when separated by two-dimensional gel electrophoresis, all differing in the number of U residues at the 3' end of the molecule. In all other cases where RNA 16 was studied only up to 5 different RNA spots were found. This difference in the number of U - residues could well be due to a poly(U) polymerase activity. Unlike most small RNA molecules discovered until now, RNA 16 does not carry any modified nucleotides. The absence of modification could not only affect its stability in the cell but also indicates that it may play a different role than for instance U1, U2 or U3 RNA. Like RNA 16, 4.5S_I RNA from Novikoff hepatoma cell nuclei (Ro - Choi et al., 1972) also does not carry any modifications, its size is approximately identical with RNA 16, it also contains a pppG.. at the 5' end and an oligo(U) stretch at the 3' end. But the nucleotide sequence of the two RNAs do not have any other feature in common. It has not directly been shown whether

RNA 16 is being synthesized by RNA polymerase III, but some indications based on the nucleotide sequence support the hypothesis of it being a polymerase III transcript. The pppG.. moiety as well as the oligo(U) stretch at the 3' end are characteristics of polymerase III transcripts. Other such examples are found in ribosomal 5S_I and 5S_{II} RNA (Ro - Choi et al., 1971). Recently Fowlkes and Shenk (1980) described an intragenic control region for the Adenovirus VA I RNA gene similar to internal promoter regions for procaryotic and eucaryotic tRNA genes (Kressmann et al., 1979; Koski et al., 1980). When RNA 16 is compared to a putative promoter "consensus" sequence for polymerase III proposed by Kressmann (doctoral thesis, 1981), some striking similarities are found.

consensus	Pu	T	Pu	G	C	Py	C	A	Pu	T	Py	Pu	Pu	...												
	*	*	*	*	*		*	*			*		*													
RNA 16	G	U	G	G	C	G	C	A	C	G	C	C	G	...												
	pos 11						23																			
...														31-45 bp	...	G	G	T	T	C	G	A	Pu	T	C	C
																	*	*	*	*	*	*	*	*	*	
...														40 bp	...	A	G	U	U	C	G	A	G	G	C	C
																pos 63								73		

The matching of the RNA 16 sequence with this putative promoter is qualitatively as good as in the case of VA I RNA or many tRNA genes. Also striking is the constancy in the location of the common sequences. The most interesting feature of RNA 16 is the fact that it carries a sequence homologous to the Alu family of interspersed repetitive sequences found at a multiplicity of a few hundred thousand copies in mammalian genomes. Even

more complete is its homology to the B₁ type sequence (Krayev et al., 1980) which is the representative of the most abundant class of mouse fold-back DNA and shows some homology to the mammalian Alu sequences. The only other discrete small RNA species which has been shown to carry an Alu sequence in eucaryotic cells is 7S cytoplasmic RNA found in virion particles of oncornaviruses (Chech and Hearst, 1975) and in uninfected avian, murine and human cells (Fowlkes and Shenk, 1980; Bell and Hardman, 1977). By Northern blotting technique Elder et al. (1981) have shown that purified, single cloned representatives of the Alu family of interspersed sequences can detect by hybridization high levels of homologous sequences in human hnRNA, the size of which ranges from about 100 - 5000 nucleotides. The size distribution of the RNA detected by Alu family probes resembles that of total RNA in poly(A⁺) and poly(A⁻) hnRNA as well as poly(A⁺) cytoplasmic RNA. However the bulk of the RNA in these cellular fractions is synthesized by RNA polymerase II. It therefore seems likely that RNA polymerase II rather than RNA polymerase III is responsible for most of the Alu specific transcription in vivo. This would mean that RNA polymerase II initiates transcription upstream from an RNA polymerase III promoter and transcribes through it. This result from Elder et al. (1981) is in agreement with the presence of Alu specific sequences in mouse snap-back RNA (Krayev et al., 1980). As shown in the Appendix, the comparison by computer analysis of the nucleotide sequence of RNA 16 with the known DNA sequence down-

stream of the globin gene allows several possible structures. In addition it is not clear what effect the tertiary interaction could have on the conformation of an imperfect hybrid composed of alternating regions of mutual homology and loop structures. In order to analyze the conformation of the DNA-RNA hybrid, three different experimental approaches were used. The first approach involved the use of uniformly labeled RNA 16 which was hybridized to cold Z-pBR322/Mchr β G-36 previously attached to Millipore filters. By subjecting the hybrid to mild treatment with nuclease T_1 , we were hoping to exclusively cleave single-stranded regions. In order to define the structure of the hybrid, both cleaved and resistant RNA fractions were subjected to fingerprint analysis. As shown in Fig. 20c, the outcome of this experiment does not completely support the hypothetical models, (namely the 5' region of the RNA shows little or no protection by the DNA). Possibly the stability of the putative hybrid structure at the 5' end of the RNA molecule is not good enough to completely protect the RNA from nuclease attack under our experimental conditions. The two other approaches were similar in the fact that they allowed an experimental study of the DNA-RNA hybrid in solution. In both experiments, terminally labeled DNA was hybridized to cold RNA and the hybrid was subjected to a treatment which should allow to differentiate between single-stranded and protected regions on the DNA. On one hand, as described in Chapter 3.7.2., the hybrid was subjected to conditions of partial degradation by S_1 nuclease, in the hope of determining the single-stranded regions by their sensitivity to S_1 nuclease. In the other experiment described in Chapter 3.7.3.,

we were hoping to differentiate between single-stranded and double-stranded regions of the hybrid molecule with the help of bisulfite, known to deaminate C residues into U residues in single-stranded DNA, followed by sequence analysis. The experiment with S_1 nuclease supported the putative hybrid models to some extent. As shown in Fig. 22, the end of the hybrid is marked by a heavy digestion of the neighboring nucleotides. The small loops of the hybrid structure however, cannot be visualized by this method. The cleavage position in the large loop is in agreement with the hypothetical model III. The cleavage site close to the labeled end of the DNA strand rather speaks for the hypothetical model V. As discussed in chapter 3.7.3., the effect of the bisulfite treatment (3rd approach) was not as expected. Nevertheless, we were able to differentiate to some extent between single-stranded and base-paired regions. The intensity of the C band in position 246 speaks against model I. Since only the C residues allow to determine the difference between double-stranded and single-stranded regions, that method, even if it was optimized, would be somewhat inadequate for hybrid molecules with short base-paired regions alternating with small loops (e.g. the small loop at position 248 - 252 on the DNA cannot be visualized because of the lack of C residues in these positions).

4.2. Intracellular distribution of small RNA molecules and speculations on their possible role

Most non-ribosomal and non-tRNA small RNA molecules of higher organisms have been found in the nuclear fraction of the cell (Zieve and Penman, 1976). For this reason they were called small nuclear RNAs. Since then, several species have been described which appear in the cell cytoplasm (Eliceiri, 1974; Walker et al., 1974). The cellular location of a small RNA species might give some indications about its role in the cell. U3 RNA for instance, has been shown to be associated with the nucleolar fraction. A portion of U3 was found hydrogen-bonded to 32 - 28S ribosomal RNA intermediates (Prestayko, 1970), which strongly suggests its involvement in nucleolar processing. When ribosomal RNA synthesis is inhibited with actinomycin D, or its processing is disturbed by labelling cells with 5-fluorouridine (Wilkinson and Pitot, 1973), U3 RNA is only found in the cytoplasmic fraction of the cell. Therefore, it seems that the retention of U3 RNA in the nucleolus is dependent upon nucleolar functioning. U1 RNA is the most abundant of the small RNA species found in the nuclei of mammalian cells. It only appears in the nucleoplasm after a lag period. A clear precursor-product relationship can be shown between RNA U1 and a rapidly labelled cytoplasmic species first described by Eliceiri (1974). This transient appearance of a nuclear RNA species in the cytoplasm is not completely understood. The same phenomenon was shown for U2 RNA. One possible explanation

could be the following: RNA U1 and U2 each carry methylated sugars at their 5' termini (Ro - Choi and Busch, 1974), which are normally only found on cytoplasmic messengers (Perry et al., 1975). It is therefore possible that the enzyme producing this modification is only found in the cytoplasm, which would be compatible with the transient appearance of these molecules in the cytoplasm. Lerner and Steitz (1979) have shown that RNA U₁ as well as a group of other RNAs from the U series can be immunoprecipitated in form of snRNPs by antibodies from systemic lupus erythematosus patients. The presence of U1 in the nucleus as well as the striking complementarity between its 5' region and sequences which span splice junctions in nuclear RNA suggests a possible role of small RNAs located in the nucleus in the splicing of mRNA precursors. Goldstein and Ko (1974) have shown the existence in amoeba proteus of small RNAs which shuttle between the nuclei of artificially constructed binucleate cells. The role of these migrating RNAs has not yet been established. 7S RNA (Weiner, 1980) which is also complementary to the Alu family sequence was found exclusively in the cytoplasm of uninfected Hela cells or associated with virions of mammalian oncornaviruses (Review by Bishop, 1978). In murine and human cells 7S RNA was reported to be associated with polysomes (Walker et al., 1974), but other groups were not able to reproduce this result (Zieve and Penman, 1976). RNA 16 was found in Friend cells in the nuclear as well as in the cytoplasmic fraction. As mentioned in the Results section,

the validity of this finding is debatable because of possible artefacts during the separation of the cellular components. However Harada and Ikawa (1979) recovered RNA 16 from nuclear and cytoplasmic poly (A) RNA of oncornavirus infected and non-infected rodent cells. Such an association could suggest that this small RNA might be involved in splicing and joining of nuclear poly(A) RNA, transport of mature messenger from the nucleus to the cytoplasm or even protection of poly(A) RNA from nuclease attack. Harada and Kato (1980) suggest a possible hybridisation of the purine-rich section of RNA 16 to pyrimidine-rich stretches in the 3' non-coding region or to intron sequences of different messenger precursors from rat and mouse. It is very difficult to evaluate the possible significance of this suggestion. First of all, the regions of homology are in most cases quite short. On the other hand, the apparent universality of this homology (globin, immunoglobulin, insulin) could reflect a very basic mode of action of this small RNA in premessenger processing. In addition, similarly as for U1 RNA, RNA 16 has some homology with splice junctions of a variety of premessengers (Harada and Kato, 1980). However, the length of the putative base-paired region is at the limit of being significant. In contrast to U1 RNA, the region of homology to precursor splice junctions are located in the 3' terminal region of RNA 16. The actual state of knowledge allows a series of speculations, but none of the models established by various groups is supported by strong data. In any case, one should remember that the bulk of cellular RNA is associated in vivo with proteins which probably play

an important role in directing their biological activity.

4.3. Association of RNA 16 with high molecular-weight poly(A) RNA.

As shown in Fig. 10, RNA 16 was recovered from high molecular weight poly(A) RNA by thermal release. Similarly, Jelinek and Leinwand (1978) have shown that a group of small RNAs can be hydrogen-bonded to nuclear and cytoplasmic poly(A) RNA from cultured Chinese hamster ovary cells. However, it is not certain whether the small RNAs are associated with poly(A) RNA in vivo. As mentioned in the Results section, this association in vitro could be due to an experimental artefact, for instance during phenol extraction. However, since it was shown that hnRNA contains Alu sequences at a very high frequency it is certain that poly(A) RNA contains nucleotide sequences complementary to small RNA molecules belonging to the Alu family. Brunel et al. (1981) have recently analyzed the structure of snRNP particles and their mode of association with hnRNPs in HeLa cells. If the hypothesis of an RNA - RNA interaction as postulated by Jelinek is correct, it seems rather astonishing that Brunel can separate snRNPs from hnRNPs in caesium chloride, under conditions where pairing interactions between hnRNA and snRNA should have been stabilized. This new result suggests that in vivo the association between small and high molecular weight RNA is mediated by protein components, and that the finding that hn poly(A) RNA binds small RNA molecules in vitro might be due to an experimental artefact. Thus, even if the molecular arrangement of the snRNP-hnRNP complexes was known, this would still not necessarily explain the function of such a complex.

4.4. Association of low molecular weight RNA with RNP particles

As shown in Fig. 11 there is some indication that RNA 16 could be present in RNP particles. Deimel et al. (1977) demonstrated the presence of low molecular weight RNA having S values of approximately 4.5 to 6.4S in RNP particles of rat liver nuclei. They postulated that the small RNAs serve as structural components for the attachment of the proteins to form monomer RNP structures. More recently the use of antibodies occurring in sera from patients suffering from autoimmune diseases has allowed the immunochemical preparation of snRNPs starting from crude nuclear extracts (Lerner and Steitz, 1979). A series of the snRNAs contained in snRNP-antigens have been characterized by Lerner et al. (1980). Anti SM, anti-RNP, anti-Ro and anti-La are 4 different groups of sera from different patients, each of which precipitates a specific set of antigenic RNPs. Anti-RNP antibodies from lupus sera selectively precipitate snRNP species containing U1a and U1b RNA, anti-Sm antibodies bind these two RNPs and four others, namely those containing U2, U4, U5 and U6 RNA, anti-Ro antibodies precipitate a different set of RNA molecules located in the cytoplasm of mammalian cells and finally anti-La precipitates a heterogeneous collection of nuclear RNAs. According to Lerner et al. (1981), RNA 16 is one of the small RNA species which can be precipitated with anti-La antibodies. The same antibodies were shown to precipitate two small RNA molecules encoded by Epstein-Barr virus. The authors have some evidence that the two RNAs reveal

a significant homology with the most conserved nucleotide sequence of the Alu family. It is reasonable to assume, that both the RNA and the protein moiety in the snRNP are involved in whatever role the complex plays in the cell. In this context it is interesting to note that RNase P of E.coli, an enzyme involved in tRNA processing, has the structure of an RNP, the RNA moiety of which is required for the enzyme activity (Kole and Altman, 1979).

4.5. Relationship between RNA 16 and the other RNA species hybridizing to Z-pBR322/Mchr8G-36 DNA

RNA 16 constitutes the most abundant non-globin RNA species in Friend cells which can be selectively hybridized to Z-pBR322/Mchr8G-36 DNA under our experimental conditions. However, the experiment in Fig. 2 shows clearly the presence of slower migrating, diffuse RNA bands which indicate the existence of higher molecular weight RNA species with sequence homology to the globin gene insert. A possible explanation for this result is the presence of a precursor-product relationship between the RNA molecules obtained from the different bands. The fingerprint analysis of the RNA species labeled RNA 5S - 1 to 5S - 6 in Fig. 4 speaks against this possibility, since some spots present in the smallest species are absent in the larger molecules. On the other hand, when the fingerprints of molecules 5S - 2 to 5S - 5 are compared (the RNA species which do not correspond to RNA 16), a certain similarity in the patterns can be noted, although it should be pointed out that a cross-contamination of the different species (as shown by the background smear in the gel of Fig. 2) might cause problems in the interpretation of

the data. Knowing the frequency of the Alu sequences in mammalian genomes it seems more likely that the different RNA species are related to the globin insert by the presence of an Alu sequence (Fig. 19a). The DNA sequences belonging to the Alu family are known to possess efficient transcriptional initiation sites for RNA polymerase III in vitro (Jelinek et al., 1980). It is not known how well these Alu sequences function as polymerase III transcription units in vivo, but it does not seem impossible that besides RNA 16 some other groups of RNA molecules smaller than 10S would carry Alu sequences in Friend cells. One other small RNA species, 7S RNA from uninfected Hela cells (Weiner, 1980), also originally identified as a component of both avian and murine oncornavirus particles, (Erikson et al., 1973) has been reported to be related to the Alu family of sequences. The low hybridization yield of the larger molecular weight species in comparison to RNA 16 can be explained either by a lower frequency of such RNAs in the cell or it could be attributed to the fact that their homology to Z-pBR322/Mchr β G-36 DNA is not as good as in the case of RNA 16. A similar case was recently described by Denisson et al. (1981), who demonstrated the presence of genomic loci for U1, U2 and U3 RNA in Hela cells containing up to 16% mismatches with the known RNA sequence. These "pseudogenes" seem to be present in the cell at a higher number than the actual genes themselves. The same situation could be postulated for this group of small RNAs from Friend cells. In fact, one could imagine a family of genes, of which Z-pBR322/Mchr β G-36 DNA would contain one representative. For some reason the Z-pBR322/Mchr β G-36 DNA copy, if one assumes that

it is able to code for an RNA molecule, would not be as active in mouse cells as the gene coding for RNA 16. The extent of the activity of a gene like Z-pBR322/Mchr6G-36 DNA could be different in various cell types or could be dependent on the developmental stage of the organism. The large difference in the length (as judged by electrophoretic mobility) of the RNA molecules however, speaks against such a theory.

4.6. What is the role of Alu sequences in the eucaryotic genome?

A large portion of the DNA from most eucaryotic organisms is organized with alternating regions of unique and repetitive sequences (Davidson et al., 1975). About one third of these interspersed repetitive sequences exist as inverted repeats (Deininger and Schmid, 1976). Also heterogeneous nuclear RNA, which is a direct transcript of the genomic DNA, contains interspersed repeated sequences, some of which can exist as intramolecular base-paired regions with loops of variable sizes (Federoff et al., 1977). It was shown by Houck et al. (1979) upon renaturation of human cell DNA to moderate Cot values followed by S_1 nuclease digestion that approximately 8% of the genome is made of interspersed repetitive sequences and that the Alu family could account for about one half of the major interspersion pattern of the human genome. This observation implies that most DNA clones prepared from genomes of mammalian cells will contain such sequences. Recently Bell et al. (1980) have observed an Alu family member located about 6kb from the 3' end of the human insulin gene. Duncan et al. (1979) have found middle repeti-

tive sequences belonging to the Alu family preceding the β - and γ -globin gene cluster in humans. When this DNA is used in vitro as template for RNA polymerase III, it produces discrete RNA transcripts with the same polarity as the globin genes. Shen and Maniatis (1980) have obtained evidence for similar arrangements of middle repetitive sequences in the β -like globin gene cluster of rabbit. In their analysis of genetic loci for U1, U2 and U3 RNA from Hela cells, Denisson et al. (1981) found two truncated divergent gene copies for U1 and U2 sequences. In each of these clones an element of the Alu family is located upstream from the snRNA pseudogene and is in the same orientation as the gene itself. In the mouse chromosomal globin clone used here the Alu sequence is located downstream of the globin gene on the same strand and in the same orientation. Kramerov et al. (1979) have cloned the most abundant class of mouse fold-back RNA (ds - RNA B). The repetitive sequences of these clones could be grouped into two non-cross-hybridizing subclasses designated B₁ and B₂. The nucleotide sequence of 3 clones of the B₁ family was established. The mismatch between the different members of this family is approximately 8%. As shown in Fig. 19b, the portion of Z-pBR322/Mchr8G-36 DNA which hybridizes RNA 16 could be one representative of these murine B₁ sequences. Upon comparison of the human Alu sequences with the mouse B₁ sequence (Elder et al., 1981), it appears that the human Alu family resembles in many features a duplication of the 130 bp long B₁ sequence (Fig. 19b). An additional characteristic of some Alu family members is the presence of short direct repeats

flanking the sequence on either side. In the case of the Alu sequence 6 kb 3' from the human insulin gene, Bell et al. (1980) have observed a perfect direct repeat of 19 nucleotides. Elder et al. (1981) have found a perfect direct repeat of 10 nucleotides flanking one of their polymerase III transcription units. In another human clone Pan et al. (1981) described a 17 base pair repeat containing one mismatch. The direct repeats themselves are not homologous from one Alu member to the other, suggesting that they arose from duplication of sequences flanking the Alu family. Almost all Alu sequences start at the 5' end with GGCT... and carry at the 3' end an oligo(A) stretch. In the case of the B₁ sequence this constancy in the end sequences has not been observed. The role of these interspersed repetitive sequences has not yet been elucidated. The fact that they have retained extensive homologies between mammalian species argues that they are functional in some way in the cell. The presence of direct repeats on either side of the Alu family sequences is reminiscent of transposable elements. However it was shown that the length of the direct repeats for any given transposable element is always the same and the length of the Alu sequence is much shorter than any known transposon. A mechanism of transposition seems attractive as an explanation for the interspersion of a repetitive sequence and would in some way explain why middle repetitive sequences have been more conserved during evolution than unique DNA sequences not coding for a protein. For instance, the propagation of such a sequence

in the genome could be directed by a mechanistic selective pressure. Only the sequences which have retained a certain primary structure will be able to further spread throughout the genome, whereas sequences which have undergone too many changes will not reproduce themselves anymore and eventually be changed to such an extent, that it will have a random sequence. Hinegardner (1976) and several other authors have suggested that the DNA of higher organisms consists of a minority of sequences with highly specific functions plus a majority with little or none. Of course, this statement is somewhat premature, since some of the DNA with no apparent function might some day be attributed a very specific role for the organism. However, since there is no direct evidence that repetitive sequences like the Alu family make any specific contribution to the phenotype of their host, the possibility of them being selfish DNA (i.e. DNA without any function) cannot completely be excluded. On the other hand, it seems a waste for an organism to conserve and replicate such a large amount of useless sequences. One possible involvement of a repetitive sequence could be related to the chromosome structure. It seems unlikely however, that Alu-like sequences would be involved in a basic mechanism like nucleosome assembly, due to the lack of periodicity in their distribution on the DNA. In addition, the nucleosome repeat length ranges from about 160 base-pairs to 241 base-pairs depending on the organism. If involved in chromosome structures, repetitive sequences of the type of Alu sequences should be

involved in a higher order interaction. Another way to explain the presence of these sequences at such a high multiplicity in the genome, would be to consider them as a pool of DNA which at some point of evolution might be used by the organism to increase its evolutionary adaptability by trying out new genes or constructing new regulatory systems. Another theory would be a possible implication of interspersed repetitive sequences in gene regulation events. Fritsch et al. (1980) have shown that in the human globin-like gene cluster, one repeat is located 5' to the δ -gene and another one 3' to the β -gene. They are located in regions thought to be involved in the cis-acting suppression of fetal globin gene in adults. Davidson and Britten (1973) have postulated that the role of repetitive sequences could be to bind diffusible regulatory molecules which could serve for the transcriptional activation of the contiguous structural genes. The similarity of Alu sequences with transposable elements supports the possibility of a role in genetic regulation. For instance, such elements could jump from one place to the other in the genome and by their presence or absence turn genes on or off. Furthermore, since interspersed sequences are separated from each other by single copy genes and DNA sequences of a total length of approximately 5000 nucleotides, they might be involved in crossing over events leading to new combinations of allelic variants. Maybe they could play a similar role as the repetitive sequences in class switch recombination regions of immunoglobulin heavy chain genes

(Kataoka et al., 1981). The replication origins of the double-stranded papovaviruses BK, SV40 and polyoma resemble a 14bp highly conserved region of the Alu family sequence. This could suggest that the Alu family may also serve as origin for host chromosomal replication (Jelinek et al., 1980). If we assume that Alu sequences could serve as origins of replication, small RNA molecules carrying Alu sequences could play the role of primers for replication. All these possible explanations for the presence of Alu sequences in mammalian genomes do not explain its presence in hnRNA. Holmes and Bonner (1974) have suggested that they might be involved in posttranscriptional processing of hnRNA by directing its secondary structure in order to expose certain sites for association with proteins or making them accessible to processing enzymes. One explanation for the high conservation throughout evolution would be that they had to act as recognition or binding sites for specific nucleases. Most polysomal messenger RNAs with known coding functions have been thought to consist predominantly of single copy sequences. Just recently Tashima et al. (1981) demonstrated that Alu sequences hybridize to cDNA made from cytoplasmic polysomal messenger RNA. It is not known however, where on the RNA the complementary sequence is located and whether this finding has any biological significance. By computer analysis, it was shown that the globin messenger RNA carries a few short regions of sequence complementarity with the Alu sequence specific portion of RNA 16 (at the maximum 13 bp at the beginning of the second

exon). The maintenance of a high degree of homology between individual members of interspersed repetitive sequence families (Alu, B₁, etc.) is remarkable and has to be explained. As shown above, the structure and distribution of these interspersed repetitive sequences allows many speculations about their possible function in the cell, but currently a final answer cannot be given.

REFERENCES

- Aloni, Y., Dhar, R., Laub, O., Horowitz, M. and Khoury, G.
(1977) Proc. Natl. Acad. Sci. USA 74, 3686-3690.
- Artman, M. and Roth, J. (1971) J. Mol. Biol. 60, 291-301.
- Attardi, G., Parnas, H., Hwang, M. and Attardi, B. (1966) J.
Mol. Biol. 20, 145-182.
- Baralle, F., Shoulders, C., Goodbourn, S., Jeffreys, A. and
Proudfoot, J. (1980) Nucleic Acids Res. 8, 4393-4404.
- Bastos, R. and Aviv, H. (1977) Cell 11, 641-650.
- Bell, A. and Hardman, N. (1977) Nucleic Acids Res. 4, 247-268.
- Bell, G. Pictet, R. and Rutter, W. (1980) Nucleic Acids Res. 8.
4091-4109.
- Berget, S., Moore, C. and Sharp, P. (1977) Proc. Natl. Acad.
Sci. USA 74, 3171-3175.
- Bergstrand, A., Eliasson, N., Hammarsten, E., Norberg, B.,
Reichard, P. and von Ubisch, H. (1948) Cold Spring Harb.
Symp. Quant. Biol. 13, 22-25.
- Bishop, J. (1978) Ann. Rev. Biochem. 47, 35-88.
- Bonner, J. and Widholm, J. (1967) Proc. Natl. Acad. Sci. USA
57, 1379-1385.
- Bonner, W. and Laskey, R. (1974) Europ. J. Biochem. 46, 83-88.
- Brack, C. and Tonegawa, S. (1977) Proc. Natl. Acad. Sci. USA
74, 5652-5656.

- Branlant, C., Krol, A. and Ebel, J.-P. (1981) *Nucleic Acids Res.* 9, 841-858.
- Breathnach, R., Mandel, J.L. and Chambon, P. (1977) *Nature* 270, 314-319.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4853-4857.
- Brenner, S., Jacob, F. and Meselson, M. (1961) *Nature* 190, 576-581.
- Brownlee, G. (1972) "Determination of sequences in RNA" North Holland, Amsterdam.
- Brunel, C., Sri Widada, J., Lelay, M.-N., Jeanteur, P. and Liqutard, J.-P. (1981) *Nucleic Acids Res.* 9, 815-830.
- Celma, M., Dhar, R., Pan, J. and Weissman, S. (1977) *Nucleic Acids Res.* 4, 4392-4395.
- Chech, T. and Hearst, J. (1975) *Cell* 15, 429-446.
- Currier, T. and Nester, E. (1976) *Anal. Biochem.* 76, 431-441.
- Curtis, P. and Weissmann, C. (1976) *J. Mol. Biol.* 106, 1061-1075.
- Curtis, P., Mantei, N., van den Berg, J. and Weissmann, C. (1977) *Proc. Natl. Acad. Sci. USA* 74, 3184-3188.
- Curtis, P., Mantei, N. and Weissmann, C. (1978) *Cold Spring Harb. Symp. Quant. Biol.* 42, 971-984.
- Dahlberg, J. (1968) *Nature* 220, 548-552.
- Darnell, J. (1978) *Science* 202, 1257-1260.

Davidson, E. and Britten, R. (1973) Quant. Rev. Biol. 48, 565-613.

Davidson, E., Galav, G., Angerer, R. and Britten. R. (1975) Chromosoma 51, 253-289.

Deimel, B., Louis, C. and Sekeris, C. (1977) FEBS Lett. 73, 80-84.

Deininger, P. and Schmid, C. (1976) J. Mol. Biol. 106, 773-790.

Denhardt, D. (1966) Biochem. Biophys. Res. Comm. 23, 641-646.

Denisson, R., van Arsdell, S., Bernstein, L. and Weiner, A. (1981) Proc. Natl. Acad. Sci. USA 78, 810-814.

Dierks, P., van Doyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.

Duncan, C., Biro, P., Choudary, P., Elder, J. Wang, R., Forget, B., De Riel, J. and Weissman, S. (1979) Proc. Natl. Acad. Sci. USA 76, 5095-5099.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Cell 20, 313-319.

Efstratiadis, A., Kafatos, F., and Maniatis, T. (1977) Cell 10, 571-585.

Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., De Riel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. and Proudfoot, N. (1980) Cell 21, 653-668

- Eliceiri, G. (1974) Cell 3, 11-14.
- Elder, J., Pan, J., Duncan, C. and Weissman, S. (1981) Nucleic Acids Res. 9, 1171-1189.
- Erikson, E., Erikson, R., Henry, B. and Pace, N. (1973) Virology 53, 40-46.
- Federoff, N., Wellauer, P. and Wall, R. (1977) Cell 10, 597-610.
- Ford, J. and Hsu, M. (1978) J. Virol. 28, 795-801.
- Fowlkes, D. and Shenk, T. (1980) Cell 22, 405-413.
- Fradin, A., Gruhl, H. and Feldman, H. (1975) FEBS Lett. 50, 185-189.
- Fraser, N., Nevins, J., Ziff, E. and Darnell, J. (1979) J. Mol. Biol. 129, 643-656.
- Friend, C., Patuleila, M. and Harven, E. (1965) J Natl. Cancer Inst. Monograph 22, 505-522
- Friend, C., Scher, W., Holland, J. and Sato, T. (1971) Proc. Natl. Acad. Sci. USA 68, 378-382.
- Fritsch, E., Lawn, R. and Maniatis, T. (1980) Cell 19, 959-972.
- Galas, D. and Schmitz, A. (1978) Nucleic Acids Res. 5, 3157-3170.
- Gilbert, W. (1978) Nature 271, 501.
- Glover, D. and Hogness, D. (1977) Cell 10, 167-176.

Goldberg, M. (thesis, 1979)

Goldstein, L. and Playt, W. (1955) Proc. Natl. Acad. Sci. USA, 41, 874-880.

Goldstein, L. and Ko, C. (1974) Cell 2, 259-266.

Granboulan, N. and Scherrer, K. (1969) European J. Biochem. 9, 1-20.

Grosveld, G., Koster, A. and Flavell, R. (1981) Cell 23, 573-584.

Gruss, P., Lai C.-J., Dhar, R. and Khoury, G. (1979) Proc. Natl. Acad. Sci. USA 76, 4317-4321.

Gruss, P. and Khoury, G. (1980) Nature 286, 634-639.

Gruss, P., Efstratiadis, A., Karathanasis, S., König, M. and Khoury, G. (1981) Proc. Natl. Acad. Sci. USA 78, 6091-6095.

Hagenbüchle, O. and Schibler, U. (1981) Cell 23, 451-458.

Ham, R. (1965) Proc. Natl. Acad. Sci. USA 53, 288-293.

Hamer, D. and Leder, P. (1979) Cell 18, 1299-1302.

Harada, F. and Ikawa, Y. (1979) Nucleic Acids Res. 7, 895-908.

Harada, F., Kato, N. and Hoshino, H. (1979) Nucleic Acids Res. 7, 909-917.

Harada, F. and Kato, N. (1980) Nucleic Acids Res. 8, 1273-1285.

Haynes, J., Kalb, V. Rosteck, P. and Lingrel, J. (1978) FEBS Lett. 91, 173-177.

- Hershey, A. (1954) J. Gen. Physiol. 38, 145-148.
- Heyden, H. and Zachau, H. (1971) Biochim. Biophys. Acta 232, 651-660.
- Hinegardner, R. (1976) Comp. Biochem. Physiol. B55, 367-370.
- Hiramaru, M., Uchida, T. and Egami, F. (1966) Anal. Biochem. 17, 135-142.
- Hofer, E. and Darnell, J. (1981) Cell 23, 585-593.
- Holmes, D. and Bonner, J. (1974) Proc. Natl. Acad. Sci. USA 71, 1108-1112
- Houck, C., Rinehart, F. and Schmid, C. (1979) J. Mol. Biol. 132, 289-306.
- Imaizumi, T., Diggelmann, H. and Scherrer, K. (1973) Proc. Natl. Acad. Sci. USA 70, 1122-1126.
- Jacob, F. and Monod, J. (1961) J. Mol. Biol. 3, 318-356.
- Jeffreys, A. and Flavell, R. (1977) Cell 12, 1097-1108.
- Jelinek, W. and Leinwand, L. (1978) Cell 15, 208-214.
- Jelinek, W., Toomey, T., Leinwand, L., Duncan, C., Biro, P., Choudary, P., Weissman, S., Rubin, C., Houck, C., Deininger, P. and Schmid, C. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.
- Kataoka, T., Miyata, T. and Honjo, T. (1981) Cell 23, 357-368.
- Kato, T. and Lowry, O. (1973) J. Biol. Chem. 248, 2044-2048.

- Kleinschmidt, A., Lang, D., Jacherts, D. and Zahn, R. (1962)
Biochim. Biophys. Acta. 61, 857-864.
- Klessig, D. (1977) Cell 12, 9-12.
- Kole, R. and Altman, S. (1979) Proc. Natl. Acad. Sci. USA 76,
3795-3799.
- Konkel, D., Tilghman, S. and Leder, P. (1978) Cell 15, 1125-1132.
- Koski, R., Clarkson, S., Kurjan, J., Hall, B. and Smith, M.
(1980) Cell 22, 415-425.
- Kramerov, D., Grigoryan, A., Ryskov, A. and Georgiev, G. (1979)
Nucleic Acids Res. 6, 697-713.
- Krayev, A., Kramerov, D., Skryabin, K., Ryskov, A., Bayev, A.
and Georgiev, G. (1980) Nucleic Acids Res. 8, 1201-1215.
- Kressmann, A., Hofstetter, H., Di Capua, E., Grosschedl, R. and
Birnstiel, M. (1979) Nucleic Acids Res. 7, 1749-1763.
- Kressmann, A. (thesis, 1981)
- Kwan, S., Wood, J. and Lingrel, J. (1977) Proc. Natl. Acad. Sci.
USA 74, 178-182.
- Lerner, M. and Steitz, J. (1979) Proc. Natl. Acad. Sci. USA 76,
5495-5499.
- Lerner, M., Boyle, J., Mount, S., Wolin, S. and Steitz, J. (1980)
Nature 283, 220-224.
- Lerner, M., Hardin, J. and Steitz, J. (1981) Science 211, 400-402.
- Lockard, R. and Lingrel, J. (1969) Biochem. Biophys. Res. Commun.
37, 204-212.

- Long, E. and Dawid, I. (1979) Cell 18, 1185-1196.
- Luse, D. and Roeder, R. (1980) Cell 20, 691-699.
- MacNaughton, M., Freeman, K. and Bishop, J. (1974) Cell 1, 117-125.
- Maki, R., Traunecker, A., Sakano, H., Roeder, W. and Tonegawa, S. (1980) Proc. Natl. Acad. Sci. USA 77, 2138-2142.
- Maniatis, T., Jeffreys, A. and van de Sande, H. (1975) Biochemistry 14, 3787-3794.
- Marbaix, G., Burny, A., Huez, G. and Chantrenne, H. (1966) Biochim. Biophys. Acta 114, 404-406.
- Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- Midgley, V. (1965) Biochim. Biophys. Acta 108, 340-347.
- Moore, K., Rogers, J., Hunkapiller, T. Early, A., Nottenburg, C., Weissman, I., Bazin, H., Wall, R. and Hood, L. (1981) Proc. Natl. Acad. Sci. USA 78, 1800-1804.
- Murray, V. and Holliday, R. (1979) Genet. Res. 34, 173-188.
- Nagata, S., Mantei, N. and Weissmann, C. (1980) Nature 287, 401-408.
- Nevins, J. and Darnell, J. (1978) Cell 15, 1477-1493.
- Nevins, J., Blanchard, J.-M. and Darnell, J. (1980) J. Mol. Biol. 144, 377-386.
- Nirenberg, M. and Matthaei, J. (1961) Proc. Natl. Acad. Sci. USA 47, 1588-1602.

- Ohshima, Y., Itoh, M., Okada, N. and Miyata, T. (1981) Proc. Natl. Acad. Sci. USA 78, 4471-4474.
- Pan, J., Elder, J., Duncan, C. and Weissman, S. (1981) Nucleic Acids Res. 9, 1151-1169.
- Penman, S. (1966) J. Mol. Biol. 17, 117-130.
- Perry, R. (1962) Proc. Natl. Acad. Sci. USA 48, 2179-2186.
- Perry, R. (1964) J. Natl. Cancer Inst. Monograph 14, 73-89.
- Perry, R., Srinivasan, P. and Kelley, D. (1964) Science 145, 504-507.
- Perry, R., Kelley, D., Friderici, K. and Rottman, F. (1975) Cell 6, 13-19.
- Peters, G., Harada, F., Dahlberg, J., Panet, A., Haseltine, W. and Baltimore, D. (1977) J. Virol. 21, 1031-1041.
- Prestayko, A., Tonato, M. and Busch, H. (1970) J. Mol. Biol. 47, 505 - 515.
- Raj, N., Ro-Choi, T. and Busch, H. (1975) Biochem. 14, 4380-4385.
- Randerath, E. and Randerath, K (1964) J. Chromatogr. 16, 126-129.
- Rigby, P., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251
- Ro-Choi, T., Reddy, R., Henning, D. and Busch, H. (1971) Biochem. Biophys. Res. Comm. 44, 963-972.
- Ro-Choi, T., Reddy, R., Henning, D., Takano, T., Taylor, C. and Busch, H. (1972) J. Biol. Chem. 247, 3205-3222.

- Ro-Choi, T. and Busch, H. (1974) In "The cell nucleus" Acad. Press, N.Y. 151-208.
- Ross, J. (1976) J. Mol. Biol. 106, 403-420.
- Sanger, F., Brownlee, G. and Barrell, B. (1965) J. Mol. Biol. 13, 373-398.
- Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. and Birnstiel, M. (1978) Cell 14, 655-671.
- Schamböck, A. (thesis, 1981)
- Scherberg, N. and Refetoff, S. (1973) Nature New Biol. 242, 142-145.
- Scherrer, K., Latham, H. and Darnell, J. (1963) Proc. Natl. Acad. Sci. USA 49, 240-248.
- Scherrer, K., Marcaud, L., Zajdela, F., London, I. and Gros, F. (1966) Proc. Natl. Acad. Sci. USA 56, 1571-1578.
- Scherrer, K., Spohr, G., Granboulan, N., Morel, C., Grosclaude, J. and Chezzi, C. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 539-554.
- Scherrer, K., Imaizumi-Scherrer, M.-T., Reynaud, C.-A. and Therwath, A. (1979) Mol. Biol. Rep. 5, 5-28.
- Sekeris, C. and Niessing, J. (1975) Biochem. Biophys. Res. Comm. 62, 642-650.
- Shapiro, R., Braverman, B., Louis, J. and Servis, R. (1973) J. Biol. Chem. 248, 4060-4064.

- Shen, C.-K. and Maniatis, T. (1980) Cell 19, 379-391.
- Shortle, D. and Nathans, D. (1978) Proc. Natl. Acad. Sci. USA 75, 2170-2174.
- Sinha, K., Fujimura, R. and Kaesberg, P. (1965) J. Mol. Biol. 11, 84-89.
- Smith, H. and Birnstiel, M. (1976) Nucleic Acids Res. 3, 2387-2399.
- Southern, E. (1975) J. Mol. Biol. 98, 503-517.
- Spohr, G., Mirault, M., Imaizumi, T. and Scherrer, K. (1976) Eur. J. Biochem. 62, 313-322.
- Stark, B., Kole, R., Bowman, E. and Altman, S. (1978) Proc. Natl. Acad. Sci. USA 75, 3717-3721.
- Szeszak, F. and Pihl, A. (1972) FEBS Lett. 20, 177-180.
- Tashima, M., Calabretta, B., Torelli, G., Scofield, M. and Maizel, A. (1981) Proc. Natl. Acad. Sci. USA 78, 1508-1512.
- Tiemeier, D., Tilghman, S., Polsky, F., Seidman, J., Leder, A., Edgell, M. and Leder, P. (1978) Cell 14, 237-245.
- Tinoco, I., Borer, P., Dengler, B. and Levine, M. (1973) Nature New Biol. 246, 40-41.
- Tilghman, S., Curtis, P., Tiemeier, D., Leder, P. and Weissmann, C. (1978) Proc. Natl. Acad. Sci. USA 75, 1309-1313.

Vogt, V. (1973) Eur. J. Biochem. 33, 192-200.

Volckaert, G., Min Jou, W. and Fiers, W. (1976) Anal. Biochem. 72, 433-446.

Walker, T., Pace, N., Erikson, R., Erikson, E. and Behr, F.
(1974) Proc. Natl. Acad. Sci. USA 71, 3390-3394.

Warner, J., Soeiro, R., Birnboim, C. and Darnell, J. (1966)
J. Mol. Biol. 19, 349-361.

Watts, J. and Harris, H. (1959) Biochem. J. 72, 147-153.

Weaver, R. and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-
1193.

Weiner, A. (1980) Cell 22, 209-218.

Wilkinson, D. and Pitot, H. (1973) J. Biol. Chem. 248, 63-68.

Wrede, P., Wust, R., Vournakis, J. and Rich, A. (1979) J. Biol.
Chem. 254, 9608-9616.

Zieve, G. and Penman, S. (1976) Cell 8, 19-31.

Zimmermann, S. and Sandeen, G. (1966) Anal. Biochem. 14, 269-277.

APPENDIX

Details of the sequence determination of the RNase T₁ and RNase A oligonucleotides of RNA 16 by further enzyme digestion.

a) RNase T₁ oligonucleotides

Spot No.	RNase	Product	Sequence	Yield
1	T ₂ A	G G	G	13.4
2	T ₂ A	$\begin{Bmatrix} C(1)* \\ G(1) \end{Bmatrix}$ $\begin{Bmatrix} C(1) \\ G(1) \end{Bmatrix}$	CG	0.8
3	T ₂ A	$\begin{Bmatrix} C(2) \\ G(1) \end{Bmatrix}$ $\begin{Bmatrix} C \\ G \end{Bmatrix}$	C ₂ G	1.8
4	T ₂ A	$\begin{Bmatrix} A(1) \\ G(1) \end{Bmatrix}$ AG	AG	3.7
5	T ₂ A U	$\begin{Bmatrix} C(1) \\ A(1) \\ G(1) \end{Bmatrix}$ $\begin{Bmatrix} C(1) \\ AG(1) \end{Bmatrix}$ $\begin{Bmatrix} (C,A)(1) \\ G(1) \end{Bmatrix}$	CAG	2.0
6	T ₂	$\begin{Bmatrix} C(2) \\ A(1) \\ G(1) \end{Bmatrix}$		

Spot No.	RNase	Product	Sequence	Yield
6 (cont.)	A	$\left\{ \begin{array}{l} C(3) \\ AC(1) \\ G(1) \\ AG(1) \end{array} \right.$	CACG + CCAG	1.8

U₂ and Physarum digestions were carried out but the amount of radioactivity was very low and the far migrating spots did not appear.

7	T ₂	$\left\{ \begin{array}{l} A(2) \\ G(1) \end{array} \right.$	AAG	1.0
	A	AAG		
	U	$\left\{ \begin{array}{l} AA \\ G \end{array} \right.$		
8	T ₂	$\left\{ \begin{array}{l} U(1) \\ G(1) \end{array} \right.$	UG	2.3
	A	$\left\{ \begin{array}{l} U \\ G \end{array} \right.$		
9'	T ₂	$\left\{ \begin{array}{l} C(1) \\ U(1) \\ G(1) \end{array} \right.$	CUG	1.3
	A	$\left\{ \begin{array}{l} C(1) \\ U(1) \\ G(1) \end{array} \right.$		
	Phys	$\left\{ \begin{array}{l} (C,U)(1) \\ G(1) \end{array} \right.$		
9''	T ₂	$\left\{ \begin{array}{l} A(1) \\ U(1) \\ G(1) \end{array} \right.$	UAG	2.3
	A	$\left\{ \begin{array}{l} U(1) \\ AG(1) \end{array} \right.$		
	U ₂	$\left\{ \begin{array}{l} (U,A)(1) \\ G(1) \end{array} \right.$		
10	T ₂	$\left\{ \begin{array}{l} C(2-3) \\ U(1) \\ G(1) \end{array} \right.$		

Spot No.	RNase	Product	Sequence	Yield
10 (cont.)	A	$\begin{cases} \text{C}(2) \\ \text{U}(1) \\ \text{G}(1) \end{cases}$	CCUG	1.4
	Phys.	$\begin{cases} (\text{C,C,U})(1) \\ (\text{C,U})(\text{trace}) \\ \text{G}(1) \end{cases}$		
11	T ₂	$\begin{cases} \text{A}(2) \\ \text{C}(2) \\ \text{U}(1) \\ \text{G}(1) \end{cases}$	AUCACG	1.0
	A	$\begin{cases} \text{AU}(1) \\ \text{AC}(1) \\ \text{C}(1) \\ \text{G}(1) \end{cases}$		
	U ₂	$\begin{cases} (\text{U,C,A})(1) \\ \text{CG}(1) \\ \text{A}(1) \end{cases}$		
13	T ₂	$\begin{cases} \text{U}(2) \\ \text{C}(1) \\ \text{G}(1) \end{cases}$	UUCG	1.5
	A	$\begin{cases} \text{U}(2) \\ \text{C}(1) \\ \text{G}(1) \end{cases}$		
	Phys.	$\begin{cases} \text{UU}(1) \\ (\text{C,G})(\text{trace}) \\ \text{C}(1) \\ \text{G}(1) \end{cases}$		
14	T ₂	$\begin{cases} \text{A}(1) \\ \text{U}(3) \\ \text{G}(1) \end{cases}$	AUUUG	1.5
	A	$\begin{cases} \text{AU}(1) \\ \text{U}(2) \\ \text{G}(1) \end{cases}$		
	U ₂	$\begin{cases} (\text{U,U,U,G})(1) \\ \text{A}(1) \end{cases}$		

Spot No.	RNase	Product	Sequence	Yield
14 (cont.)	Phys.	$\left\{ \begin{array}{l} (A,U,U,U)(1) \\ (A,U,U)(\text{trace}) \\ G(1) \end{array} \right.$		
15'	S ₁ Phosphatase	GTP mostly phosphate (quite faint)		0.9
15''	S ₁	GDP		
12/16'/16''/ 16''' (same sequence with increasing amount of U-residues at the 3' end of the molecule)	T ₂ A U ₂ Phys.	$\left\{ \begin{array}{l} C(1) \\ A(1) \\ U(1-3) \\ \text{(increasing in each spot)} \end{array} \right.$ $\left\{ \begin{array}{l} AU(1) \\ AC(1) \\ U(1-2) \\ C(\text{trace}) \end{array} \right.$ $\left\{ \begin{array}{l} ?(1-2) \text{ must be AC, but there was not enough radio-activity to further characterize this spot} \\ UU(1-2) \\ (U,C,A)(1) \\ A(1) \end{array} \right.$ $\left\{ \begin{array}{l} ? \\ (U,U,A) \\ (C,U) \\ (A,A,C,U) \\ (C,U,A) \\ U \end{array} \right.$	$ACUAU_{2-5}NOH$ or $CUACACAU_{2-5}NOH$	1.9

b) RNase A oligonucleotides

Spot No.	RNase	Product	Sequence	Yield
1	T ₂	U	U	4.2
	T ₁	U		
2	T ₂	cyclic U	cyclic U	3.7
	T ₁	cyclic U		
3	T ₂	{ A(1) C(1)	AC	2.8
	T ₁	AC		
3'	T ₂	C	C	2.9
	T ₁	C		
4	T ₂	{ G(1) C(1)	GC	2.7
	T ₁	{ G(1) C(1)		
5	T ₂	{ A(1) U(1)	AU	0.9
	T ₁	AU		
6	T ₂	{ G(1) A(1) C(1)	AGC	0.8
	T ₁	{ AG(1) C(1)		
7	T ₂	{ G(2) C(1)	GGC	1.3
	T ₁	{ G(2) C(1)		
10	T ₂	{ A(1) C(1) U(1)	AGU	1.2
	T ₁	{ AG(1) U(1)		

Spot No.	RNase	Product	Sequence	Yield
11	T ₂	{ G(2) C(1)	looks like GGC but position speaks absolutely for GGGC	1.0
	T ₁	{ G(2) C(1)		
12	T ₂	{ A(1) G(3) C(1)	(AG)(G) ₂ C	0.9
	T ₁	{ AG(1) G(2) C(1)		
13	T ₂	{ A(2) G(3-4) C(1)	(AG) ₂ (G)C	1.0
	T ₁	{ AG(2) G(1) C(1)		
16	T ₂	{ A(2) G(2) U(1-2)	(AG)(G)AU	1.1
	T ₁	{ AG(1) AU(1) G(1)		
17	T ₂	{ A(1) G(2) U(1)	(AG)(G)U	
	T ₁	{ AG(1) G(1) U(1)		
18	T ₂	G(2) U(1)	GGU	4.0
	T ₁	G(2) U(1)		

Spot No.	RNase	Product	Sequence	Yield
19	T ₂ T ₁	background in all 4 { runs like A ₃ C C	ppGpC	0.7
20	T ₂ T ₁	some C { runs like A ₃ U C some Pi	pppGpC	
21	T ₂ T ₁	{ A(2) G(2-3) U(1) AG(2) AU(1) G(2)	(AG) ₂ (G) ₂ AU	1.3
22	T ₂ T ₁	{ A(2) G(2) C(1) A ₂ G(1) AG(2) G(2-3) C(1)	(A ₂ G)(AG) ₁₋₂ (G) ₂₋₃ C	0.8

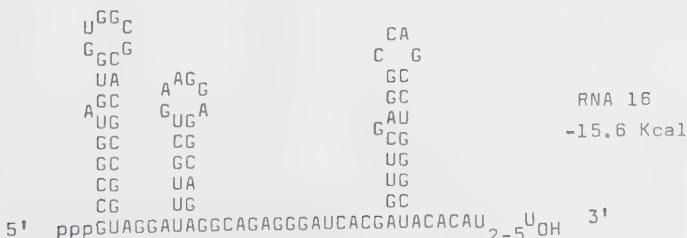
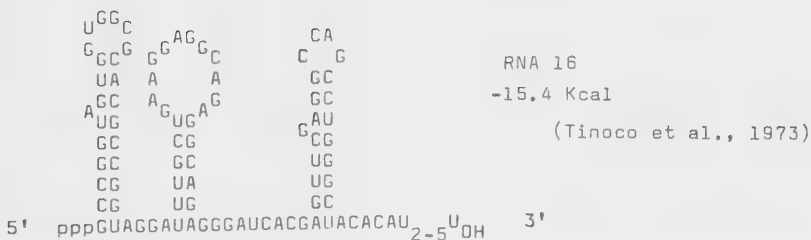
Footnotes:

As shown in Tables 2 and 3 the estimation of the quantities of certain oligonucleotides were too low. The most probable explanation is an incomplete transfer of these oligonucleotides from the cellulose acetate strip to the DEAE paper.

*The number in parenthesis indicates the molar ratio of every component.

The nucleotide composition of most oligonucleotides resulting from U₂ or Physarum RNase digestions were determined by the R_f value of the spot. In only a few cases the amount of radioactivity was high enough to allow a further analysis with T₂ RNase.

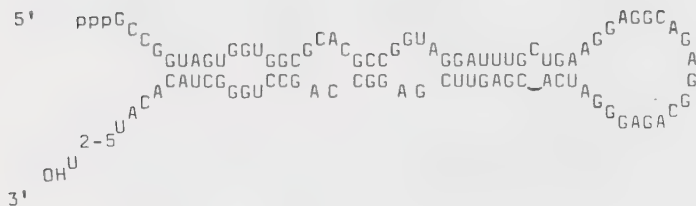
possible secondary structures



secondary structures (cont.)

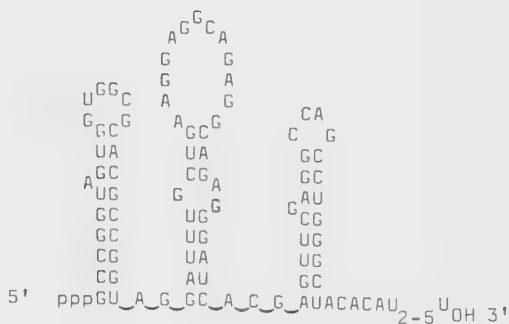
RNA 16

-11.6 Kcal



RNA 16

-11.6 Kcal



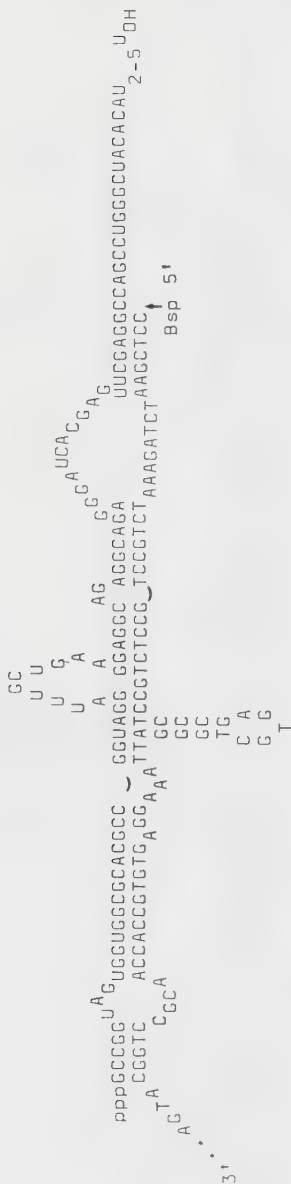
OR

—

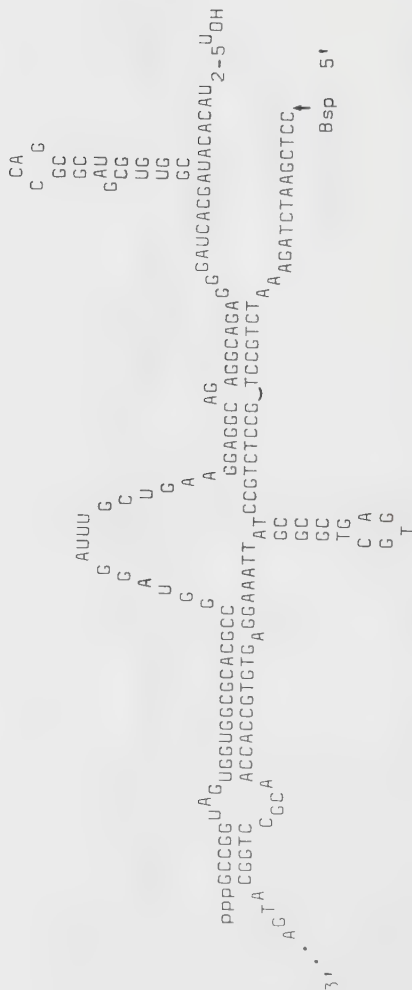
—

..T T C.
or
AT C
CC
CC
CC
TC
C A C
C T

IV



V



Curriculum Vitae

Name Grüber-Steinitz Christine Annelise

Bürgerort Vugelles-La-Mothe (VD)

Geburtsdatum 6. Feb. 1952 in Sierre (VS)

Vorbildung Primarschule 1959-1963 Heiligenschwendi (BE)
Progymnasium 1963-1967 Bern
Gymnasium 1967-1971 St.-Maurice (VS)

(1 Jahr Unterbruch von 1969-1970 für Studienaufenthalt in Saratoga, California, USA)

1971 Eidgenössische Matura Typ B

Studium Studium der Biologie an der Universität Lausanne,
WS 1971/1972 - WS 1973/1974
1972 1er examen propédeutique
1973 2e examen propédeutique
Weiterführung meines Studiums am Zoologischen Institut der Universität Zürich, SS 1974 -
WS 1976/1977
1974 2. Vorprüfung Hauptfach
1977 Schlussdiplom
(Diplomarbeit ausgeführt am Zoologischen Institut unter der Leitung von Herrn Prof. Dr. P.S. Chen)
Seit Oktober 1977 als Assistentin am Institut für Molekularbiologie I tätig.
Dissertation unter der Leitung von Herrn Prof. Dr. C. Weissmann.

Verzeichnis der
Dozenten deren
Vorlesungen ich
besucht habe
(Universität
Zürich)

Bachofen, Baker, Batschelet, Baumann, Billeter,
Birnstiel, Briegel, Burla
Chen, Christen
Dübendorfer
Gasser
Hohl, Humbel
Jungen-Hauschreck
Kägi, Kubli, Kummer
Melli
Nievergelt, Nöthiger
Pflüger
Rast
Tardent, Thomas
Weber, Wehner, Weissmann, Wüthrich
Ziswiler

